# Influence of base pair mismatch location on the binding efficiency of nucleotide strands

THESIS

submitted in partial fulfillment of the
requirements for the degree of

BACHELOR OF SCIENCE
in
PHYSICS


| | |
|---|---|
| Author : | Arjan G. van Breemen |
| Student ID : | 1357352 |
| Supervisor : | Prof. dr. Dirk Bouwmeester |
| $2^{nd}$ corrector : | Dr. Peter Gast |


Leiden, The Netherlands, July 20, 2017

# Influence of base pair mismatch location on the binding efficiency of nucleotide strands

**Arjan G. van Breemen**

Huygens-Kamerlingh Onnes Laboratory, Leiden University
P.O. Box 9500, 2300 RA Leiden, The Netherlands

July 20, 2017

## Abstract

The thermodynamic principles of complementary binding of DNA or RNA strands are well known, thus the binding efficiency can be calculated for a given sequence. However, when a two-stranded molecule contains mismatches in the base pairs, e.g. a guanine opposite to thymine base, traditional models like the nearest-neighbour model do not suffice. The influence of the location of these mismatches on the binding efficiency, in particular, is not well understood. Understanding the binding behaviour of nucleotide strands is essential to the development of applications that require efficient and highly exclusive binding to specific sequences in RNA or DNA. Such an application is exon skipping, a gene correction therapy to treat Duchenne muscular dystrophy. An alternative model has recently been developed at Leiden University to explain this mismatch location dependency. This study is aimed at comparing its calculations with experimental results obtained by DNA encapsulated silver nanocluster fluorimetry. This type of fluorimetry uses a special labelling technique to relate the emitted intensity of various bulk samples to the binding efficiency. Mismatches are introduced in various locations, and the binding efficiency is measured to determine the dependency on the mismatch location. The binding efficiency as a function of the mismatch location shows a relation that resembles the calculations by the model. Improvements of the method are suggested based on the results, allowing for a valid evaluation of the model.

To Hans Rochat, who lost his son Marc to Duchenne Muscular Dystrophy



Version of July 20, 2017– Created July 20, 2017 - 18:18

# Contents

# List of Figures

Chapter 1

# Introduction

## 1.1 Background

Duchenne muscular dystrophy (DMD) is a disease known for its slow progression and impact on a patient's life. It disables the muscles' ability to repair, hence it leads to muscle failure that spreads over time. Eventually, the heart and respiratory system lose their function, leading to an early death, usually before 30 years of age [4].

DMD is caused by hereditary defects in the genome that code for the protein dystrophin. A person with DMD produces a dysfunctional version of dystrophin. Proper dystrophin connects membranes inside muscle fibres to prevent damage during contraction, such as rips [5].

Some treatments that are currently in development use the principle of skipping the defects in the code during the process of gene expression. This leads to a short version of dystrophin that in contrast to the mutated DMD gene is still functional [Aartsma oldest and newest]. Many promising studies have been conducted using mice and cell cultures [6] and in clinical trials, the method has been moderately successful [7].

This skipping mechanism can only perform effectively when it binds efficiently and exclusively to the defect parts of the sequence. This exclusiveness is of high importance since various cell functions will be impaired when similar but "healthy" sequences have their translation into proteins altered. This requires a thorough understanding of the binding behaviour of strands, especially those with nearly identical sequences. The ther-

modynamic principles governing the hybridisation[1] of complementary[2] stands are reasonably well understood through sequence dependent models such as the nearest-neighbour (NN) model [3] [8–10]. In the case of the DMD gene therapy mentioned previously, "healthy" sequences different from the target do not match the binding site of the drug exactly. However, binding can still occur when only some single bases are different, which are called "mismatched" bases [11]. The knowledge of mismatched binding behaviour is not sufficiently comprehensive yet to prevent treatments from affecting other sites securely [6, 12], providing a challenge for current studies into nucleotide[4] thermodynamics.

In recent studies, several important factors have been identified such as the type of mismatched base pairs, the type of strand configuration, i.e. DNA/RNA, DNA/DNA or RNA/RNA[5] and the hybridisation technique: surface or solution and individual strands or bulk [13, 14]. It has been suggested that the location of a mismatch in the sequence is a dominant factor hence this influence is studied at Leiden University and in this BSc research [13, 15, 16]

## 1.2 Goal

This study aims at developing a better quantitative understanding of nucleotide binding for short duplexes with mismatches in their base pairs. In particular, the binding efficiency dependency on the location of the mismatches in the sequence is of interest. This dependency can be obtained from models, however, these still require their validity to be evaluated experimentally. Using Ag-DNA encapsulated nanocluster fluorimetry, experimental results can be compared to the relation calculated by a model developed at Leiden University in order to improve it. This is likely to

[1] Hybridisation is the binding of single strands that form a double stranded complex (a duplex), which occurs at a sufficiently low temperature

[2] Able to form a perfectly matching double-stranded structure with the opposite nucleotide strand.

[3] A model to calculate the binding energy of a strand by using sequence dependent parameters, one for each nucleobase and its adjacent base. The traditional version of this model can only be applied to perfectly matching strands and that is the only well established one.

[4] The basic building block of nucleic acids, such as DNA and RNA. It is an organic compound made up of nitrogenous base, a sugar, and a phosphate group. An elaboration can be found in *Acronyms and definitions.*

[5] Here "/" means "bound opposite to"

contribute to the development of treatments of DMD at the LUMC.

## 1.3 Approach

To obtain the binding efficiency for each mismatch location in a sequence, experiments are conducted using specifically designed DNA strands that are (near) complementary to a DNA excerpt of the DMD gene. These strands are identical, except for a single altered base that will form a mismatch with the DMD gene excerpt. The binding efficiency is determined using samples in which the location of this mismatch in the sequence is different for each sample. The binding efficiency for each strand is obtained by measuring the emission spectrum of an Ag-DNA probe. The binding efficiency is related to the emitted intensity through the strong dependence of the optical properties of the Ag-DNA probe to its DNA environment. This property is explained in Theory 2.4. Since studies into this binding behaviour commonly use aggressive dyes and synthetic binding surfaces such as microarrays. Hence the approach of this BSc research is in principle less subject to undesired effects on the binding efficiency and reflects physiological conditions better. Moreover, it is cheaper, simple and the experiments can be performed in a shorter amount of time. With these measurements, the mismatch location dependency can be graphed and compared the calculated result generated by the model.

Chapter 2

# Theory

## 2.1 Application to gene correction therapy for Duchenne muscular dystrophy

### 2.1.1 DMD and its cause

An improved understanding of the binding behaviour of DNA duplexes[1] containing mismatches[2] can substantially contribute to a wide range of medical applications. In particular, it can be applied to a gene correction therapy, that potentially provides a treatment for Duchenne muscular dystrophy (DMD). The therapy is based on targeted binding of the active drug components to specific sequences in pre-mRNA, thus requiring efficient and highly exclusive binding [17] .

DMD is a hereditary muscular disease that causes a major decrease in a patient's mobility and drastically lowers the life expectancy [18, 19]. It affects roughly 1 in 5000 live male births worldwide [20]. The muscles degrade by a deficit of functioning dystrophin, which is a protein that prevents damage to the muscular tissue. It serves as an "anchor" connecting elements moving relative to each other in the muscle: the actin cytoskeleton and the myocyte membrane [21]. This causes the muscles to lose functionality and they are replaced by fat tissue consequently. Several genetic defects such as deletions and duplications in the dystrophin gene are responsible for

[1] A DNA structure consisting of two strands bound opposite of each other

[2] A mismatch in this context means a change in the pattern of the bond locations inhibiting the formation of a base pair, e.g. a G base opposite of a T base instead of the fitting C base.

producing truncated proteins that fail to form this connection [5, 6, 22].

An example of a genetic defect that causes DMD is a mutation in a section named exon 51. An exon is a section of the pre-mRNA sequence that is processed into the mRNA. Therefore, an exon will be translated into a protein. Introns are sections of pre-mRNA that are skipped in this process, contrary to exons. This selective skipping process is called "splicing" [23]. Due to mutations in the DMD gene, exon 51 does not appear in the mRNA after splicing, leading to a frameshift in the mRNA sequence. This shift in the reading frame is not a multiple of three bases, which is the length of a codon [3]. Therefore other groups of three bases are falsely interpreted as the codons that are to be translated into proteins. Thus, this shift means that the reading is disrupted, resulting in abnormal codons being translated [24]. Consequently, one codon is misread as stop-codon and the produced dystrophin protein is truncated, thus not functional. The corresponding mRNA strand is out-of-frame. [25].

### 2.1.2 Exon skipping: a treatment for DMD

Skipping this exon introduces a further shift that makes the total shift a multiple of three. This is again in-frame since codons consist of three bases. Although causing the protein to be shorter, this will not lead to a truncated protein and therefore it is functional nevertheless. As a result, the effect of the dystrophy is significantly decreased. Instead, a form of Becker muscular dystrophy (BMD) or even the absence of dystrophy symptoms is achieved. [24, 25]. BMD is a disease similar to DMD that is far less severe than DMD, consequently, BMD patients have far higher life expectancy than those with DMD[22].

To force the skipping of normal exons such as exon 51, antisense oligonucleotides (AONs) can be used. These are synthetic single strands with nucleobases that "hide" the exon to which they bind during the splicing process [26]. Furthermore, the skipping of exon 51 can restore the reading frame in more mutation cases: 13 % of deletions can be treated by targeting this exon, more than other exons [6]. Figure 2.1 displays a schematic of an exon skipping technique from E.M Conner et al. [1]

[3] A codon is a group of three nucleobases coding for one amino acid, which is the constituent of a protein

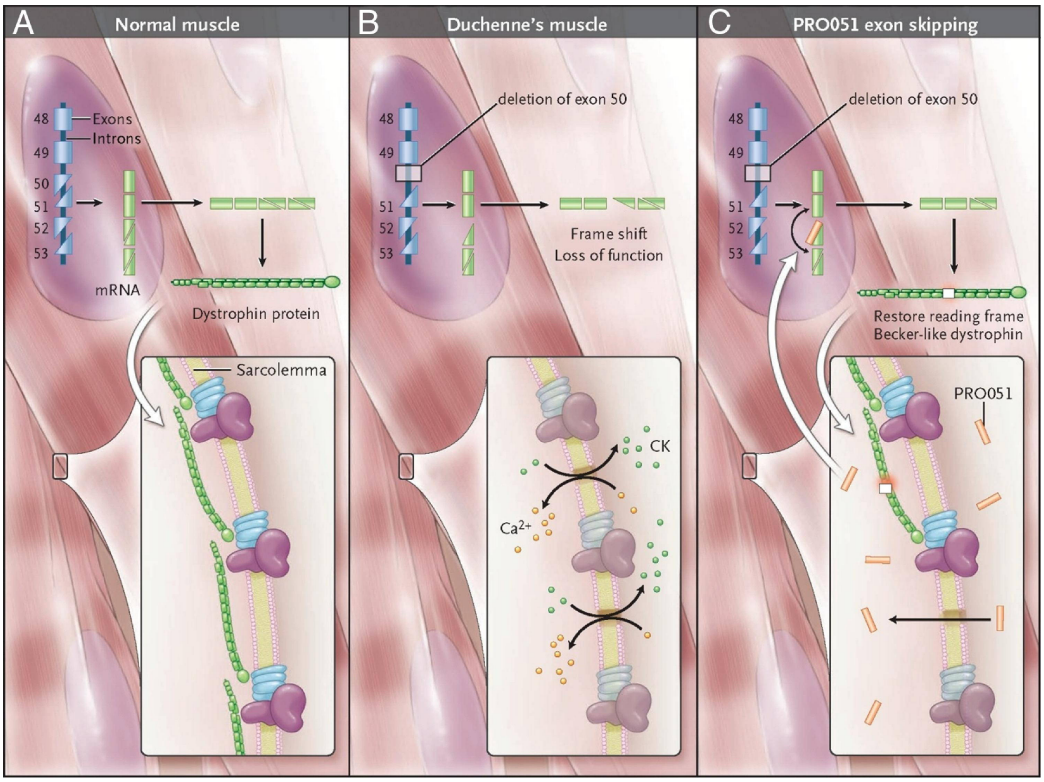


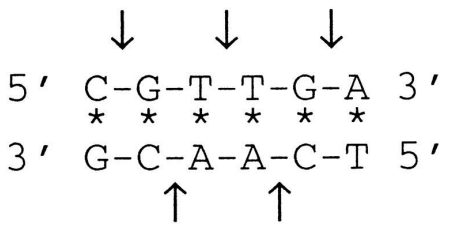

***Figure 2.1:*** *An example of the principle of the exon skipping treatment [1]*

Various other treatments for DMD are studied and applied as well. However, the advantage of exon skipping is its ability to halt the progression of the disease. The common non-genetic therapies, such as the use of steroids, only help to maintain muscle functionality for limited periods. Other approaches that instead of exon skipping or non-genetic treatments edit the DMD gene were found to be inefficient at the current stage of development. The exon skipping with AONs has been more successful as well. Additionally, gene editing methods, such as CRISPR/Cas9 permanently modify the genetic material, in contrast to exon skipping. Therefore these techniques are subject to ethical objections and safety concerns [6, 12, 27].

## 2.2 Nucleotide hybridization: the binding behaviour of complementary strands

### 2.2.1 Bonding structure

The hybridisation of nucleotide strands is a crucial function inside living cells that allows for gene expression, which is a vital mechanism in the human body. Various diseases and treatments involve altered gene expression or binding to specific sites on DNA and RNA [26]. Nucleotide strands form stable duplexes by the forces present in the hydrogen bonds between opposite nucleobases (cytosine, guanine, adenine, thymine or uracil) [28]. A hydrogen bond is a noncovalent bond between a molecule and a hydrogen atom from another molecule in which the rest of the latter is more electronegative than the hydrogen. [29]. Longer duplexes contain more bases, thus more bonds and are therefore generally stronger. Additionally, a duplex containing more guanine-cytosine (GC) pairs than one of equal length is generally stronger since a Guanine-Cytosine (GC) pair is bound by three hydrogen bonds instead of an adenine-thymine (AT) pair, which consists of only two [30]. The hybridisation can be quantified by binding strength, which is a force or thermodynamic binding quantities, i.e. free binding (Gibbs) energy, entropy and enthalpy. Additionally the melting temperature$T_M$, which is the minimum temperature at which 50% of strands are not hybridised, [31] or the binding efficiency, i.e. the percentage of strands that form a duplex when single strands are mixed in bulk are useful quantities. In this study, that last quantity is used.




$$\Delta G^\circ_{37}(\text{pred.}) = \Delta G^\circ(\text{CG/GC}) + \Delta G^\circ(\text{GT/CA}) + \Delta G^\circ(\text{TT/AA}) + \Delta G^\circ(\text{TG/AC}) + \Delta G^\circ(\text{GA/CT}) + \Delta G^\circ(\text{init.})$$

$$= -2.17 - 1.44 - 1.00 - 1.45 - 1.30 + 0.98 + 1.03$$

$$\Delta G^\circ_{37}(\text{pred.}) = -5.35 \text{ kcal/mol}$$

$$\Delta G^\circ_{37}(\text{obs.}) = -5.20 \text{ kcal/mol}$$

***Figure 2.2:*** *Nearest neighbour method: calculation of the free energy of a duplex [2]*

### 2.2.2 Nearest-neighbour method

As the binding energy of duplexes appears to be highly dependent on their structure, this induces the development of computational models to predict the binding energies using the sequence of a complementary strand.

Indeed such a model has been developed which is well-known for its accuracy: the nearest-neighbour (NN) method [32]. The principle of this method is the summation of free energy elements $\Delta G$ corresponding to segments of the strand to obtain its total free binding energy. These segments consist of a base pair and the neighbouring pair e.g. GC opposite to CG in this explanatory figure 2.2 [2]. An initial free energy element is required to account for the terminals since they have only one neighbour.

## 2.3 The binding behaviour of mismatched strands

### 2.3.1 Background

The traditional NN method, however, only applies to complementary strands and duplexes can contain mismatches, e.g. a guanine opposite of a thymine base, with the different hydrogen bond patterns inhibiting the formation of a base pair [8]. The study of the binding behaviour of these duplexes is currently at an early stage and models based on the sequence have been unable to predict the binding energies of mismatched duplexes as accurately as complementary strands.

Models, calculations or databases that have been developed to predict the effect of mismatches often aim to contribute highly in depth to the theory of the binding behaviour [33]. Although providing a thorough understanding, a more general approach is believed to improve the application with a higher rate. Said in depth studies involve mainly NN parameter corrections for mismatches [2, 8, 34–36] or the binding energies corresponding to various DNA structure configurations caused by mismatches [10, 11, 37–39]. Applications request insight into the explicit effects of mismatches on the binding efficiency of strands in bulk instead [40].

Most studies into general sequence-dependencies of the binding efficiency focus on experimental findings or models for hybridisation in or on other media than those similar to human cell environments, e.g. microarrays [33, 41, 42]. Furthermore, there are many studies aimed at the design of highly mismatch intolerant oligonucleotides or sequence detectors [17, 43–47]. These tend to rely on ”massive” machine learning, e.g. neural networks, or trial-and-error [48].

They have contributed to applications to the extent that clinical trials with AONs for DMD treatments were moderately successful [7]. Nevertheless, these programs or large databases are not suitable for proper predictions of binding efficiency due to their analytic nature and therefore do not significantly improve the (qualitative) theoretical understanding. The author of this BSc thesis believes this is essential for the AON-methods ever to lead to a real cure for DMD since a substantial prolongation of life has not yet been achieved [12].

### 2.3.2 Location dependency

The location of a mismatch in a sequence is suggested to be an important factor in the binding efficiency [14, 49, 50], although this has often been denied or neglected before [36, 38]. For a single mismatch, which means only one base alteration per strand, this dependency seems to be dominant over the type of nucleobases at a mismatch, e.g. a guanine base opposite of a thymine base [13, 15, 16]. However, there is no clear consensus on this [14, 51] as the type dependency is complex, and more sensitive to other influencing factors [52].

Generally, mismatches positioned in the centre of the sequence tend to result in a lower binding efficiency, while positions closer to the 3′ or 5′ end lead to higher efficiencies for strands roughly between 15 and 20 bases long [13–15, 53]. Other studies that focused on the specificity of binding, i.e. the tendency of a strand to not bind with mismatched targets, indicate the same result. [52, 54]. Additionally, in some of these studies, the relation was not found by the design of strands with single base changes bound to one target sequence. Instead, a method was used in which various different strands were mixed and those forming a stable configuration were categorised by their sequence to count the mismatched strands per position.

In contrast, strands of roughly 20 to 25 bases often increase in binding efficiency when the mismatch is positioned in the middle as well as close to the ends. It decreases in between, resulting in the W-shape that can clearly be seen in the result from this study: [55]. This has also been found in unpublished work available to the author of this BSc thesis: [56]. The same effect, however less clear, is mentioned in these papers: [14, 54]. Even longer sequences are less affected by a single mismatch since there are many binding sites available to form stable bonds, making them less sensitive to the position of the mismatch [54].

Most studies attribute these position relations to effects caused by the DNA being surface-bound, which indeed has an influence on the binding efficiency [8, 14, 52]. Furthermore, the labels that are used to distinguish bound strands from unbound ones can affect the binding as well, especially aggressive dyes [45, 57, 58]. However, the efficiency increase close to the ends has been observed or suggested in solution based research as well [49, 51, 56]. These apparent similarities suggest that these studies are useful in physiological context, although the methods used can cause a

strong deviation from in vivo binding behaviour. They can direct the aim of this research topic, nevertheless repeating them with dye-less solution based experiments is required.

Models or theories that explain the effect of mismatches on the binding efficiency commonly do not incorporate the location dependency explicitly. Moreover, there is little consensus between them and they are not considered generally valid [59]. Additionally many fail to explain or predict the increased binding efficiency at central mismatch positions that appears in longer strands, mentioned previously.

The approach of location dependent models varies, e.g. a position dependent correction on the NN model [41, 60] or principal statistical thermodynamics such as Langmuir [8, 13, 15, 53] or Sips [58, 61]. Some of these models are purely based on experimental data, e.g. those introducing new NN parameters, which may be problematic to generalise for any sequence or environment. This, in combination with the focus on arrays and influence of dyes on the results on which these models are based, requests the development of novel models.

### 2.3.3 The model used in this thesis

A model aiming to predict or explain the explicit effects on binding efficiency as a function of the position of a single mismatch in a strand has recently been developed at Leiden University. The work has not been published yet (July 2017) and the author of this BSc thesis provides the necessary information about the model in this work on the basis of communication with its developers and this article draft: [56].

To avoid this complexity of effects on the secondary structure of the DNA, the model used in this study treats the mismatches not as nucleotide elements, but as "gaps" in the strand instead. With this approach, the strand is virtually divided into two sections: one on the left of the mismatch and one on the right. These separate sequences are now essentially complementary, allowing one to apply the traditional NN model, which is more accurate than mismatch models [9]. The complex behaviour of the mismatches is then introduced in a simplified way through a correction in the energies, representing the average effect of a mismatch on DNA binding. This allows for the explicit calculation the predicted binding efficiencies

for any sequence to determine the generalised behaviour. This correction includes the dominant influence of ”dangling ends”.

A dangling end means that a part of the strand does not form hydrogen bonds with the target strand, while the rest of the strand does. The duplex is partially ”unzipped” and one part ” dangles loosely” under Brownian motion [37]. The longer the dangling part, the weaker the duplex and in this model, its effect is linear with the mass of the dangling section. As explained earlier, the longer a (section of) a strand, the more bonds can be formed, and the more stable the duplex is. Therefore both statements indicate that a sequence section can either be too short to bind or too long to prevent unzipping, leading to a nonintuitive balance that explains the complexity of the relationship that has been observed. In this model, the position of the mismatch governs the various dangling end configurations leading to large differences in binding efficiency for different positions.

The model calculates the total enthalpy change $\Delta H$ using two terms; one corresponding to the sequence to the left of the mismatch and the other to the one on the right. This limits the applicability of this calculation to non-terminal mismatches, i.e. not at either end of the sequence, since only one of the two halves exists when the mismatch is at a terminus. Each of these enthalpy terms contains a term determined by the traditional NN method minus a correction for the effect from the concatenation of the mismatch and the rest of the sequence. In an equivalent manner, the total entropy change $\Delta S$ is calculated. The total entropy and enthalpy changes are used to predict the binding efficiency for a mismatched sequence. When this is calculated for all positions, the position dependency can be obtained in a format that is fitted to experimental results. The developers of the model describe the used equations as follows:

“The changes in enthalpy ($\Delta$H) and entropy ($\Delta$S) upon the binding of two DNA strands of length N can be determined from:

$$\Delta H = \Delta h_0 + \sum_i^N \Delta h_i \tag{2.1}$$

$$\Delta S = \Delta s_0 + \sum_i^N \Delta s_i \tag{2.2}$$

using the previously determined energy changes from an individual base

pair $(\Delta h_i, \Delta s_i)$ in the nearest neighbour model [9, 10][4]. The melting temperature, $T_m$, can then be determined from

$$T_m = \frac{\Delta H}{\Delta S + R\ln(C)} + 16.6\log(Na^+) \tag{2.3}$$

where C is the total concentration of oligonucleotides when both strands are in equal numbers, and [Na+] the molar concentration of monovalent cations in the solution." The binding efficiency introduced as $\theta$ is expressed in the following manner:

$$\theta = 1 - \frac{2}{1 + \sqrt{1 + 8\exp(-\chi)}} \tag{2.4}$$

with

$$\chi = \frac{\Delta H}{R} \cdot (\frac{1}{T} - \frac{1}{T_M}) \tag{2.5}$$

"

The enthalpy and entropy terms corresponding to the sections of the sequence left and right from the mismatch are

$$\Delta H'_{left/right} = \Delta H_{left/right} - J_H(1 - \theta_{left/right}) \tag{2.6}$$

and

$$\Delta S'_{left/right} = \Delta S_{left/right} - J_S(1 - \theta_{left/right}) \tag{2.7}$$

where $J_H$ and $J_S$ are the corrections for the separate sections being connected with a mismatch. The binding probabilities of the separate sections, derived from NN equations, are $\theta'_{left}$ and $\theta'_{right}$. These are used as weights for the energy terms in eq 2.6 and 2.7 in the calculation of the total energy changes:

$$\Delta H = \Delta H'_{left}\theta'_{left} + \Delta H'_{right}\theta'_{right} \tag{2.8}$$

and

$$\Delta S = \Delta S'_{left}\theta'_{left} + \Delta S'_{right}\theta'_{right} \tag{2.9}$$

[4]The original reference numbers have been replaced by the ones corresponding to the bibliography in this BSc thesis

None of the models described in 2.3 are well-established and consequently request various refinements. This requires effective and short experiments to test their validity. An appropriate method that serves this purpose is Ag-DNA encapsulated fluorimetry.

## 2.4 Ag-DNA encapsulated nanocluster fluorimetry

When silver atoms are clustered on the nanoscale and excited by a monochromatic beam, they display a fluorescence with plasmonic properties [62]. These clusters quickly lose this property due to agglomeration into bulk objects. This is prevented by encapsulating the strands in molecular structures like DNA. As a consequence, these molecules will determine the shape of the cluster and therefore the optical properties that depend on the aspect ratio of the nanocluster [62, 63]. When DNA is used, the fluorescence is photostable [64] and highly sequence (length) dependent [63–69].

Even a single base change in an encapsulating DNA structure shifts the emission spectrum significantly, indicating a very high sensitivity to the DNA structure [3, 70] which can be seen in figure 2.3. This sequence specificity provides opportunities for effective targeting of short ssDNA (single strand DNA) [63] and is already applied in some studies [70–73].

Several encapsulation probes for silver with clear and distinct optical properties have been found in the last 20 years, however, the sequence dependence is not well understood yet. Notwithstanding, Machine Learning methods are promising tools to improve the insight into this behaviour [74] and this publication attempts to explain the effect [75]. It has been found that the clusters favour the adherence to G or C bases in general [66, 76] and the emission spectrum shifts to higher wavelength range with increasing sequence length [62, 69].

Furthermore, certain encapsulations cause the otherwise biocompatible silver suitable for in vivo applications. This toxicity is even tuneable, thus targeted medical applications are possible as well [63].

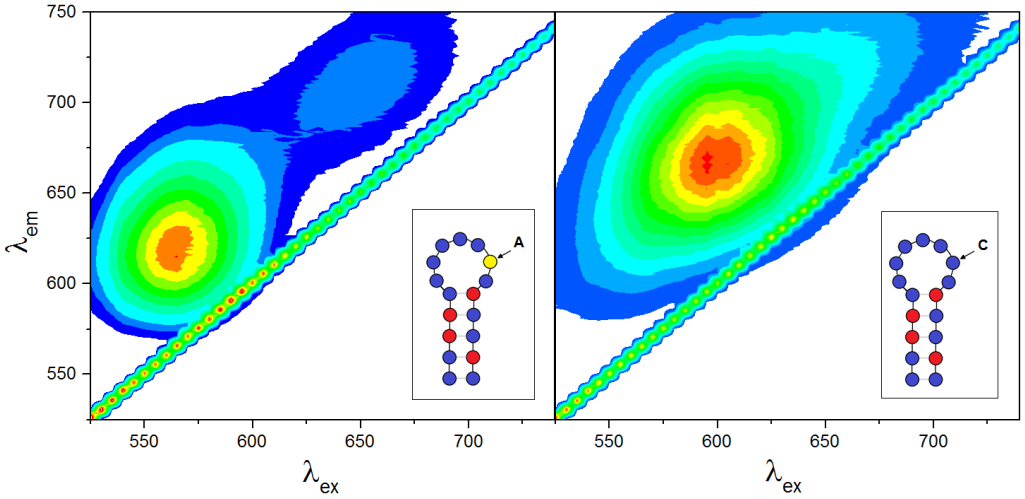



***Figure 2.3:*** *Heat map of intensity for each excitation wavelength (horizontal) and emission wavelength (vertical). When an adenine (A) base in the encapsulating nucleotide structure (left) is altered to a cytosine (C) base (right), the fluorescence changes significantly. The diagonal "line" is caused by scattering of the excitation frequency into the emission measurement.[3]*

Chapter 3

# Methods

## 3.1 General

### 3.1.1 Probe-AON technique

This study used an Ag-DNA fluorescence technique to study the binding sensitivity of mismatched strands. This requires a fluorescence behaviour from the Ag-DNA that relates in a specific manner to the binding efficiency, i.e. the intensity of the emission increases with increasing binding efficiency. Not all strands display this relation , therefore a specific sequence is selected during the AON-testing experiments that clearly distinguishes the unbound and bound configuration through the measured emitted intensity.

The used strands are concatenations of an Ag encapsulation strand and a stand identical to an antisense oligonucleotide (AON) sequence. The Ag encapsulation strand, called 19b-probe, is a short sequence that has been shown to display a bright emission of the silver clusters [63, 69], which is illustrated in the schematic in figure 3.1 and figure 3.2.The blue dashed line in figure 3.2 corresponds to this strong fluorescence. In this study and previous ones, it is found that its intensity is significantly lowered when concatenated with AON-sequences [56] as seen in figure 3.1, with the corresponding spectrum as the black line in figure 3.2. Any fluorescence from other sources will have only low intensity in the wavelength domain that is measured in this experiment, as explained in Theory 2.4.

The AON-sequences are designed to bind to a target sequence: a 90 base

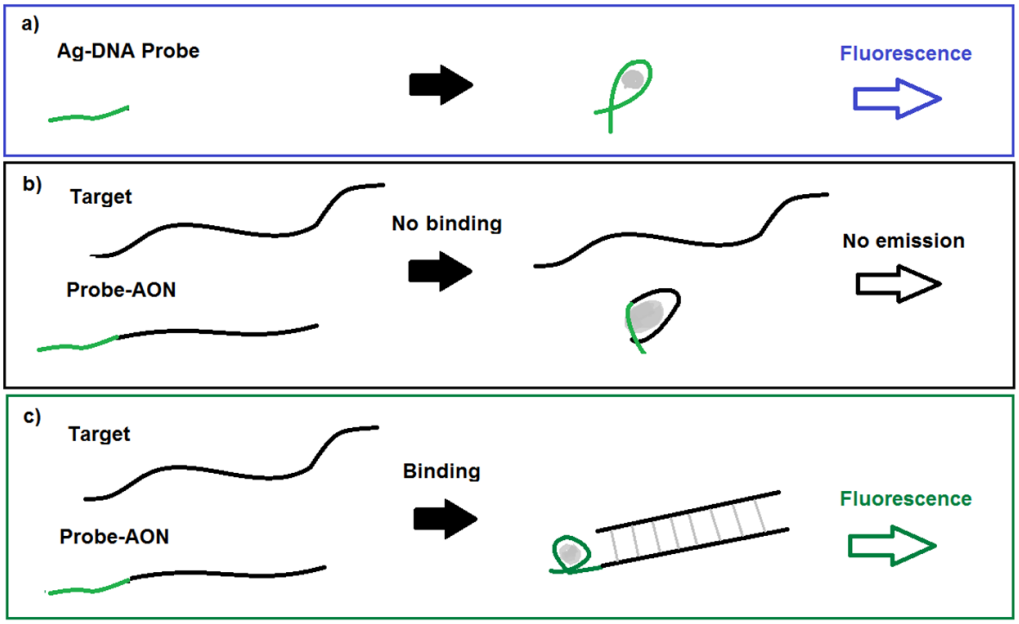



***Figure 3.1:*** *Cartoon of Probe-AON technique to measure binding efficiency through fluorescence intensity. The corresponding emission spectra can be seen in figure 3.2. The AON measured here is not used in this research.*

excerpt of exon no.51 of the pre-mRNA sequence[1] that produces the protein dystrophin when expressed. This exon is of interest in research into gene correction therapy for DMD [26], including studies in Leiden at the LUMC. An elaboration on exons and AONs can be found in Theory 2.1.2. When this target strand binds to the probe-AON strand, the AON part of the strand cannot encapsulate the silver cluster, as its bases are already bound to the targeted exon sequence. This leaves only the probe part "free" to influence the optical properties of the cluster, resulting in a nearly identical spectrum as emitted by the probe alone as seen in figures 3.13.2 (green line).

Using Ag-DNA in a solution to obtain the binding efficiency, a substantial influence on the result is prevented that would be present when the more common fluorescent dyes were used. In contrast to microarray studies, the method is cheap and the experiments are short and simple. The manufacturing of the custom strands is done by an external company, and from design to receiving the DNA took no longer than a week. In one session,

[1] The sequence can be found in 8.2

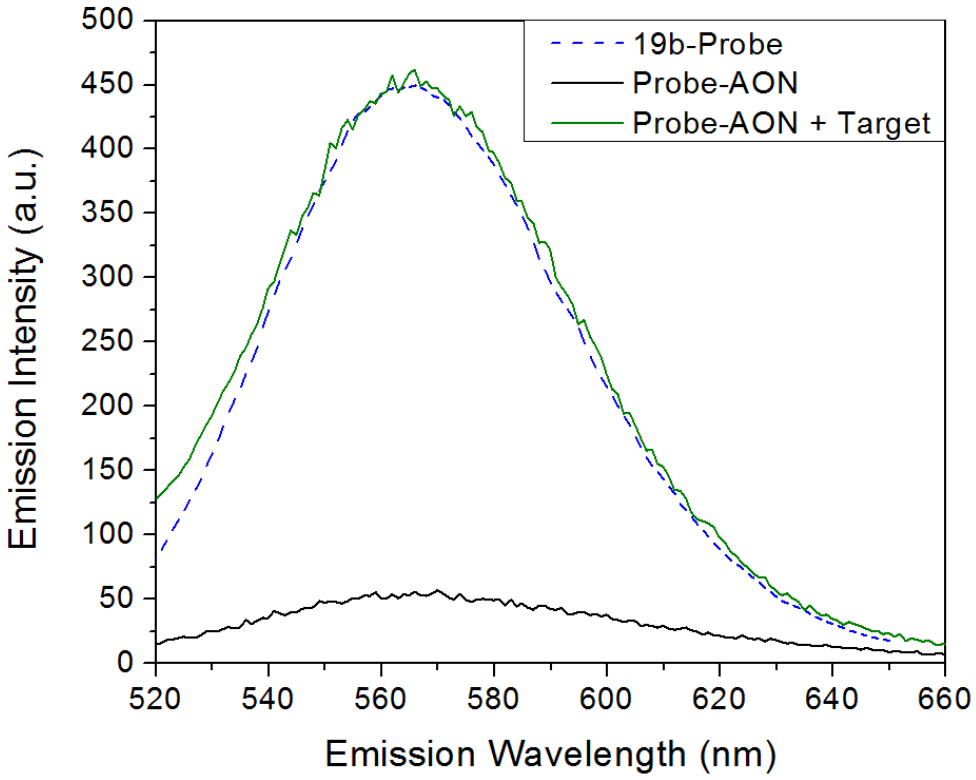



***Figure 3.2:*** *Emission spectra corresponding to the Ag-DNA probe method in figure 3.1.*

roughly 10 samples can be prepared and measured, which takes roughly 6 hours for the author of this thesis and for experienced laboratory workers at least a few hours due to waiting time. Six sessions are likely to produce a sufficiently accurate result.

## 3.1.2 Procedures

### 3.1.2.1 Preparing samples

To measure the fluorescence intensities, the strands are sampled in bulk according to the recipe specified in the Appendix: 8.3. The recipe for the samples is based on what is known to produce a strong fluorescence from the 19b-Probe. The solutions and samples are made in 1.5 mL microcentrifuge tubes using variable volume single channel pipettes. All chemicals are dissolved in nuclease-free water to protect the DNA material. Sample preparation and measurement are performed in a HEPES-NaOH, pH 7.4 buffer, representing conditions fairly close to physiological conditions within the cell. The tubes are kept at 37 $^{\circ}C$ to resemble intracellular fluid and, after inserting all DNA, given 2 hours for hybridization to take place. Afterwards, $AgNO_3$ (99.9999%, Sigma Aldrich), containing the silver atoms, is added, followed by $NaBH_4$ (99%, Sigma Aldrich), which neutralises the silver ions. During those last steps, the samples could not be kept at the 37 $^{\circ}C$, but samples were at room temperature for no longer than 30 minutes during this research. After synthesis, the fluorescent samples are placed inside a Cary Eclipse fluorimeter (Varian) to obtain their emission spectra.

### 3.1.2.2 Measurement

The samples were kept at 37 °C Celsius in the dark for half an hour between the addition of the reduction agent and measurement during the AON-testing experiment and an hour during some of the mismatch experiment sessions. A Peltier-element was used to keep the cuvette inside of the fluorimeter at 37 °C as well. With the fluorimeter, the emission spectrum of each sample is measured at an excitation wavelength of 495 nm, representing the excitation maximum of the used 19b-Probe [69]. In this research, the emission was measured from 500 nm to 650 nm, with a step size of 1 nm.

#### 3.1.2.3 Data processing

The resulting curves of the samples can be integrated over all emission wavelengths to obtain their total emission. The emission spectrum is near-Gaussian, the wavelength step size is small and only the intensity of the samples relative to each other is relevant in this research. Therefore the total emission is obtained by summing over the emission spectrum, to approximate the integral as a Riemann sum, which simplifies data processing. Due to scattering of the excitation from the measurement device, for wavelengths close to the excitation, the spectra do not display the emission accurately. Therefore, the summation starts at 565 nm instead, the known emission maximum of the 19b-probe. The symmetry of the near-Gaussian shape allows the higher wavelength-half of the spectrum to represent the entire spectrum.

## 3.2 AON-testing

The purpose of this experiment is selecting one AON-sequence from 5 test sequences with an emitted intensity that increases with binding efficiency, which is required in the next stage of this study. For this purpose, the emission intensity of several complementary duplexes of the sequence is compared to their single strand configuration. Since in the first case, the strands bind optimally and the latter cannot bind because there is no complementary strand, a maximum difference in intensity is expected. The sequences are referred to as Test-AON1 to Test-AON5. Although their name suggests that these are only AON-sequences, all of them contain the 19b probe sequence on the 5' end followed by the actual AON-sequence, so the 3' end will always be "dangling" to encapsulate the silver cluster. Also, note that these are not actual AONs, but DNA strands with identical nucleobase sequences, unlike AON strands which typically consist of modified RNA. Two sets of five samples are produced for the five test sequences. In one of those sets, the complementary strand, the exon sequence, is added to accomplish the bound configuration. To the other one, the same volume, this time of nuclease-free water, is added instead of the complementary DNA solution. Consequently, no DNA-binding occurs in those tubes. For each sample, the summed intensity of the test sequence with exon is compared to the same sequence with only water. The ratio between these can be calculated for each AON in order to select the AON with maximum intensity in bound configuration and minimum intensity

in unbound configuration. In many cases, this can already be concluded from the emission spectra by eye.

## 3.3 Mismatch experiment

With the AON-sequence selected based on the results from AON-testing, another set of measurement sessions is performed. The purpose of these experiments is to obtain the mismatch location dependency of the binding efficiency. This is achieved by altering one base of the AON-sequence per sample and varying the position of this base change. In doing so, samples are produced for each possible location of the altered base, thus the set of AON/Exon[2] complexes is mismatched on 20 positions in total. Each sample, thus each mismatch location, is measured to get the intensity per location. The exact base changes can be found in the sequence table in the Appendix 8.2. As explained in the Discussion of the AON-testing: 5.2, AON4 was chosen for the mismatch experiments. During most of the sessions, the order of measurement and addition steps during the sample preparation are randomised to minimise time-dependent influences on the result.

In order to compare the results from different sessions and to relate the intensity to the binding efficiency, the intensities are normalised by the (average) intensity at the end of the strand, i.e. location 20 or 19 in sessions during which location 20 is not measured. Ideally, one would normalise with the complementary strand, however, this was not done because those results proved unreliable, as explained further in the Discussion. In the final result, all intensities are normalised by the average of location 20 and are averaged per mismatch. The standard deviation corresponding to the data per mismatch location is also obtained from these normalised intensities, to represent the error in the measurement.

## 3.4 Fitting model with parameters

The prediction by the model developed at Leiden University described in the Theory, 2.3.3, is fitted to the experimental line of the mismatch loca-

[2]The slash means "bound to" in biochemical convention, so "A/B" would be strand "A" forming base pairs with the opposite strand "B"

tion relation. There are two parameters to set corresponding to the $J_H$ and $J_S$ terms in equations 2.6 and 2.7. These are set manually for the best fit under the restriction that they are small relative to $\Delta S$ and $\Delta H$, allowing for the linear corrections $J_H$ and $J_S$ to be appropriate. In these results, they represent less than 1% of the total energetic change, meaning this is the case.

## 3.5 Control experiment: determining error in sample preparation

Before an accurate result can be obtained from the AON-testing and the mismatch experiment, the error due to the inconsistent preparation of samples, e.g. imprecise pipetting, should be limited. To determine this error, five samples of only the 19b-probe, or GS4T, are measurement. Because no AONs and target strands are included, the volume of buffer used is 80 $\mu l$ and the 20 $\mu l$ DNA solution is added. Like before, the intensities are obtained through a Riemann sum and the standard deviation divided by the average intensity is calculated to determine the error.

## 3.6 Control experiment: checking fluorescence stability

Since problems arose considering the precision of the measurement, the stability of the intensity from several samples is evaluated. During the time required for the silver to be fully neutralised by the reductor, the emission spectrum from one sample is obtained every 30 seconds. Like before, each spectrum half is summed to obtain the intensity per time interval and the intensity through time is graphed. The result is fitted with a double exponential to check whether the curve levels to a constant value, which would indicate stability.

Chapter 4

# Results

The raw data from the fluorimeter contains the measured intensity in arbitrary units (a.u) for each emission wavelength in nm, with a step size of 1 nm.

## 4.1 AON-testing

### 4.1.1 Measurements from the first session

To identify which AON-sequences would be suitable to use in our mismatch experiments, during two sessions the test-AONs were measured. As explained in Methods 3.2, for each test-AON the spectrum is measured in bound and unbound configuration, to compare these states. In the bound configuration, the binding of the AON-sequence to the 90b-target is efficient and the intensity of the spectrum should be high. In the unbound configuration, the target is replaced by nuclease-free water, so the entire test-AON can influence the silver fluorescence. From the AON design, a far lower intensity is desired and expected.

In figure 4.1 the spectra from the test-AONs, measured during the first measurement session, are graphed. The intensity of test-AON5 with target exon (red solid line) is clearly higher than without (red dashed), with the unbound sample producing a background fluorescence less than 3% of the signal when binding occurs. For test-AON4, this holds as well, although the difference between the bound configuration (black solid line) and unbound (black dashed line) is smaller than for AON5, the background fluorescence from the unbound sample being about 14% of the

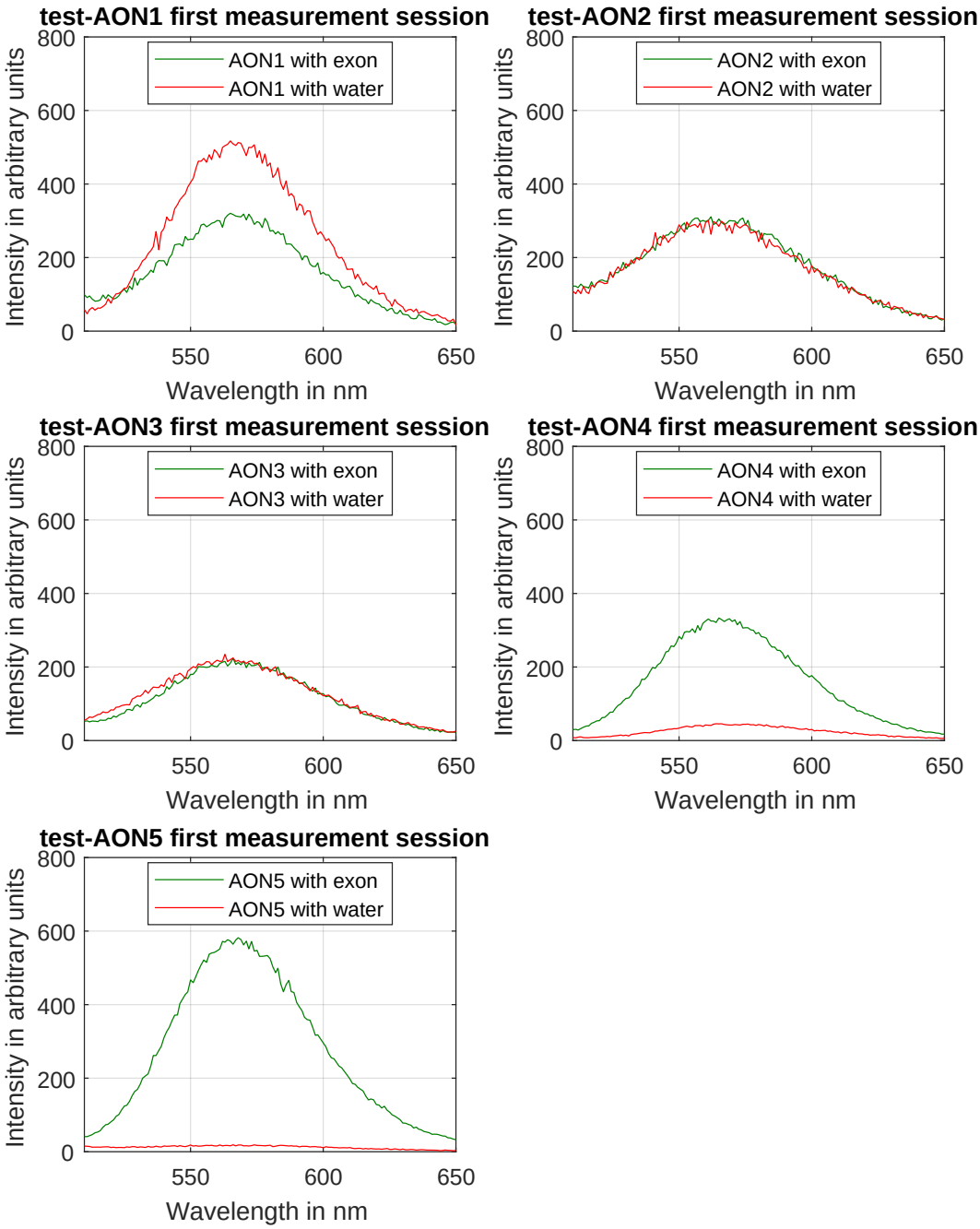



***Figure 4.1:*** *The emission spectra from the probe-AON strands, measured with the added exon target in one sample (bound configuration) and only nuclear-free water added to the other (unbound). These were measured during the first session.*

highest measured signal. AON2 and AON3 show a negligible difference between their two configurations: they overlap for many wavelengths. AON1 even shows an opposite behaviour: the intensity in bound configuration (cyan solid line) is clearly lower than in unbound configuration (cyan dashed line). All spectra are approximately Gaussian and are peaked at roughly 565 nm, in agreement with the expected fluorescence of the 19b-probe. The peaks from the two lowest curves: unbound AON4 and AON5 [1]are slightly shifted, and seem to have a peak at roughly 570 nm.

### 4.1.2 Measurements from the second session

To confirm our results, the experiment was repeated, as shown in figure 4.2. Generally, the same relations between the unbound and bound states can be seen in this repetition. An exception being that test-AON2 also shows a higher emission intensity in the unbound state (dashed deep blue line), than when it is bound (deep blue line). Again the curves are near-Gaussian and have the same peak wavelengths as those from the first session.

[1]Although challenging to see for AON5 in this figure

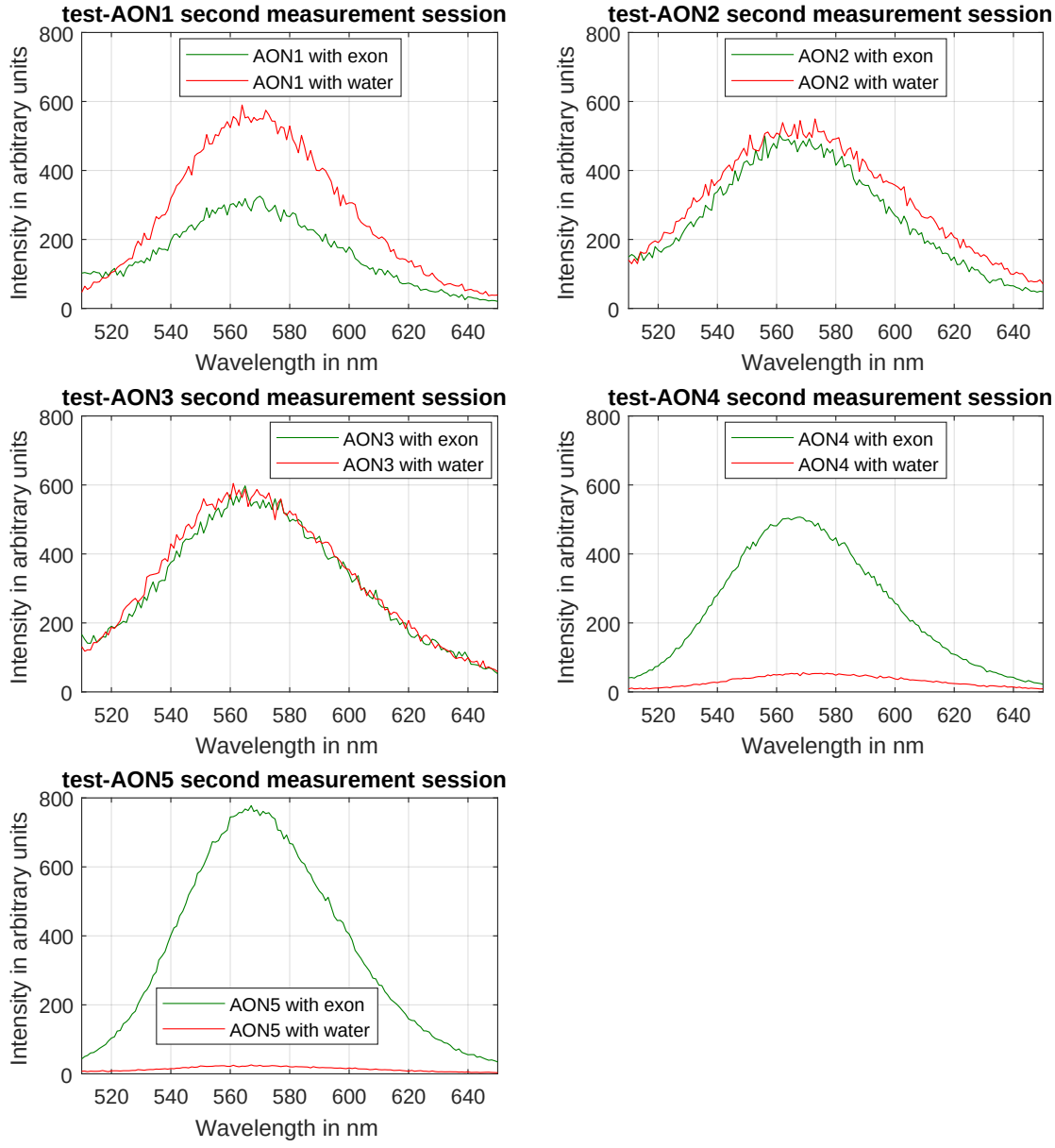



***Figure 4.2:*** *The emission spectra from the probe-AON strands, measured with the added exon target in one sample (bound configuration) and only nuclear-free water added to the other (unbound). These were measured during the second session.*

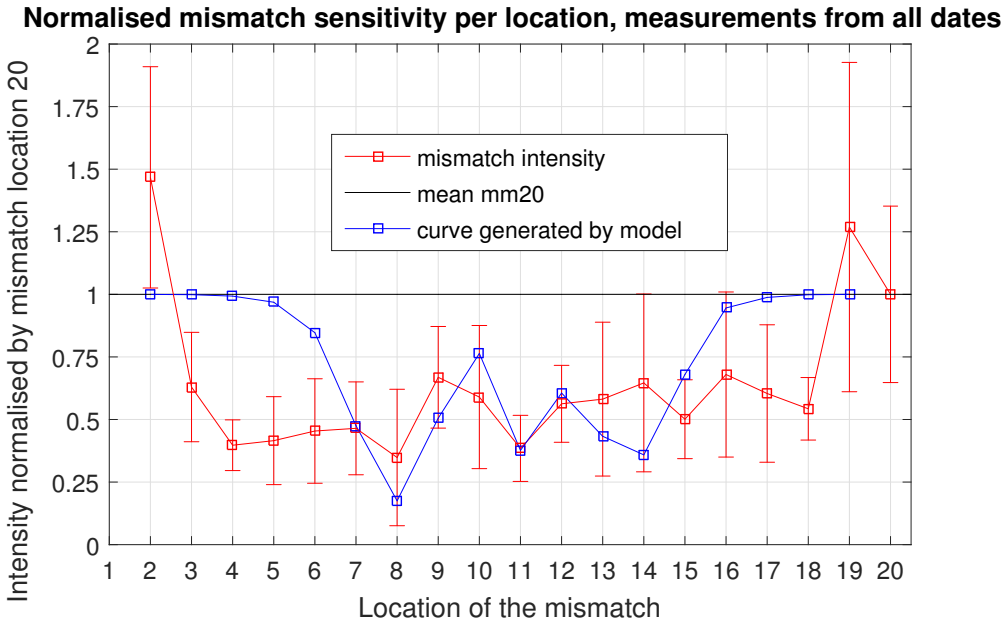



***Figure 4.3:*** *The sensitivity of binding efficiency to the location of the single mismatch expressed as the related intensity. This is normalised by the value of mismatch (mm) position 20, here plotted as the black line. In red, the averaged experimental result and in blue, the model fit.*

## 4.2 Mismatch experiments

Single-base mismatches were introduced into the sequence in all possible locations, and the fluorescence intensities were measured after hybridization and silver cluster synthesis. In doing so, we measured the mismatch dependency of the hybridization of our AON sequence, test-AON4. In figure 4.3 the model prediction is plotted in blue with the experimentally found line in red. The normalised intensities represent the total emission from the samples, obtained from the Riemann summation of the spectra. The model predictions are the result of the calculations presented in Theory 2.3.3. Using the Ag-DNA probe, the measured fluorescence intensity is positively related to the number of bound strands, allowing us to use it as a measure of the binding efficiency. The lines between points are plotted to simplify the interpretation, however, the results are discrete, there is no intensity or binding efficiency between the integers of the mismatch position.

The intensities were normalised by the average intensity of mismatch 20

for reasons explained in the Discussion 5.3. For clarity, this intensity is plotted as a black line. The points on the red line are averages of the normalised data from 9 measurement sessions. The error bars depict the standard deviation of the data sets. The binding efficiency for a mismatch at position 1 is not calculated by the model, and produces unpredictable results due to the proximity of the mismatch to the silver cluster, and is therefore excluded from the measurement.

For most locations, the measured intensities are below the normalisation meaning the mismatch affects the binding of the strands negatively, as expected. A notable exception is location 2, which is likely to also be the result of the proximity of the mismatch to the silver cluster. When approaching the terminals, the calculated curve increases to 1 much more rapidly than measured behaviour. In the middle range, in particular, between locations 7 and 12, both model and experimental results display similar peaks and dips. Most average intensities from the experimental data are around 0.5, with a maximum at position 2 of approximately 1.5 and a minimum at position 8 of approximately 0.35. At the terminals, the error is especially large, roughly 0.7 for positions 2 and 19 and the error is not less than 0.2 for any location.

The model is the result of exact calculation and does not indicate uncertainties. Its predicted binding efficiencies range from approximately 0.17 at position 8 to 1 at position 19.

## 4.3 Control experiment: checking fluorescence stability

From several measurement sessions, the intensity over time is graphed in figure 4.4 to determine if it has been a constant at the start of the session. The measured behaviour is characterised by an exponential function which stabilises to a constant after a certain time.

In this figure, it is clear that the 19b-probe (green) displays this behaviour, showing fair stability after approximately 30 minutes. However, when the AON is added to the probe sequence (black), or when including a mismatch (blue), stability was not reached within anywhere close to the same timeframe. Both intensities from the probe-AON strand and mismatched



Version of July 20, 2017– Created July 20, 2017 - 18:18

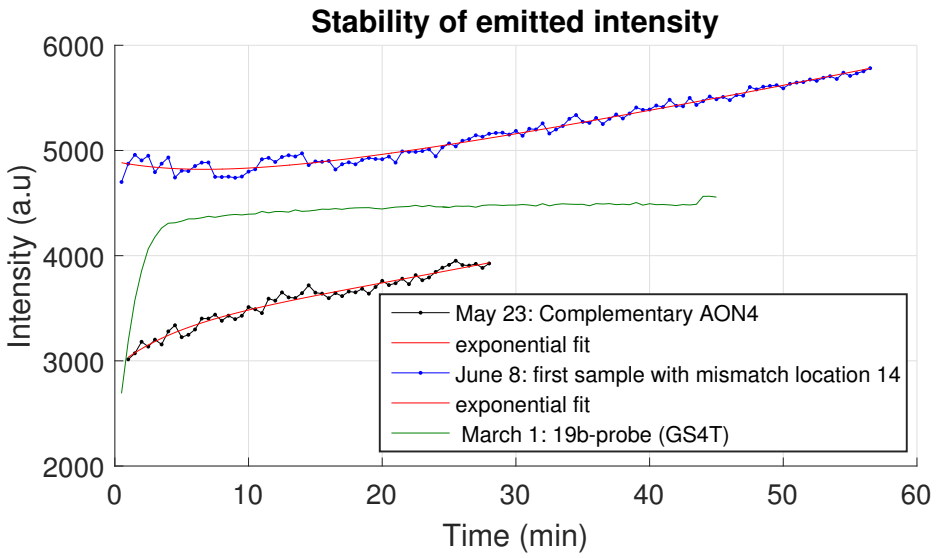



***Figure 4.4:*** *(Rescaled) intensity development over time from three samples and the exponential fit to these curves.*

strand are not constant during the time of the measurement, respectively from 28 minutes on and 56 minutes on, and the behaviour is clearly different for the two cases as well. The curves have been rescaled in the vertical direction to be able to compare them in one view, which preserves their development.

## 4.4 Control experiment: determining error in sample preparation

In figure 4.5 the spectra are graphed from five samples containing the 19b-probe that were prepared with identical content and conditions to determine the human error involved in their preparation. The overall intensity of these spectra is calculated and the standard deviation of these numbers is determined. The standard deviation normalised by the average of the intensity was found to be 13.37%, thus the error corresponding to human imperfections in sample preparation is approximately 13%.

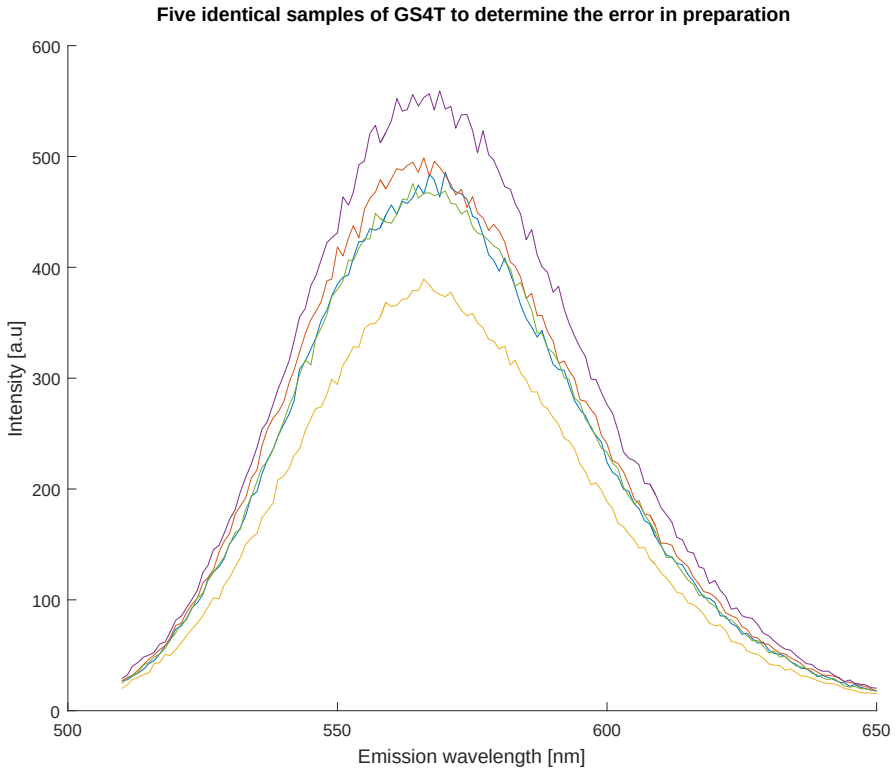



***Figure 4.5:*** *The spectra from five samples containing the prepared 19b-probe GS4T prepared according to the same recipe. From high to low: a purple, an orange, a blue, a green and a yellow curve.*

Chapter 5

# Discussion

## 5.1 Inaccuracy in the Ag-DNA probe method

When bases from the target bind to the AON-strand, the resulting structure can have various configurations that vary in binding energy. This means that ”bound” and ”unbound” are not the only states of hybridisation [38]. The Ag-DNA probe method is used to relate emitted intensity to the binding efficiency, however, it can fail to distinguish configurations that are only partially bound, i.e. ones with ”dangling ends”. This term is explained in Theory 2.3.3.

Dangling ends can occur on one of the sides next to the mismatch and if the probe-end, i.e. the 5′ end, is dangling, it is likely to affect the encapsulation in a manner that lowers the emission in the measured wavelength range. This means that, in particular when a large part of the strand close to the cluster is dangling, it will appear as ’unbound’ in the measurement, even though it is partially bound. Especially when the mismatch is close to or at the 5′ end, the effect of a short dangling part at this end is highly unpredictable. Furthermore, the 3′ end may dangle in various degrees, whilst not affecting the encapsulation on the 5′ end, leading to similar emitted intensities that fail to distinguish these various weak bonds from a non-dangling 3′ end.

Consequently, the probe-AON method is less sensitive to binding weaknesses on the 3′ side of the mismatch, when, as in our experiments, the probe is attached to the 5′ end. Therefore, it is recommended for future studies to design strands with the probe on the 3′ end as well.



Version of July 20, 2017– Created July 20, 2017 - 18:18

In the context of the application of actual AONs, considering duplexes with dangling ends as bound is less problematic. This is reasonable since exon skipping as described in Theory 2.1.2 still functions when only part of the AON is connected to its target. Thus it can be debated whether the results found in this study corresponds even better to the effectiveness of AONs in treatments than to the binding efficiency of the entire strand.

## 5.2 AON-testing

### 5.2.1 Evaluation of usability

From the spectra in figures **??** and **??**, it is clear that test-AON4 and test-AON5 cause the probe to emit a far higher intensity when bound to their complementary parts of the target exon, which is required to accurately relate the binding efficiency to a measured intensity. A reverse relation, which AON1 displays, is not linear since the low-intensity region is noisy, contrary to the positive relation. Therefore the method is more accurate when AON4 or AON5 is used.

### 5.2.2 Hypothesis on spectral differences

The difference between the AON-designs' spectra can be attributed solely to the sequences since all samples were prepared and measured under the same conditions. The author of this thesis suggests that the encapsulation affinities of AONs 1,2 and 3 were lower. Thus this resulted in the ordinary probe encapsulation even without the exon, and the dangling AON-parts of the strands did not interfere. Generally, the number of G or C bases and the length are a factor in this affinity [62, 66, 69, 76], however, based on the used sequences in 8.2, neither of these could have caused the differences.

It is not plausible that the problems considering the fluorescence instability affected these results. The instability is treated in another section in this chapter. Contrary to the mismatch experiments only the intensity difference between two samples is important. These are the one with the target and the one without, that were measured in quick succession. [1] This is why there is no instability expected during the AON-testing sessions.

[1] within 5 minutes

The spectra from AON4 and AON5 without target show a very low intensity since a fairly homogeneous distribution of configurations is formed when the 19b-probe stabilises fluorescent clusters. The measurement is tailored to this expected probe signal. The measured signals without the target are thus mostly the background from a small number of 19b-probe encapsulations forming naturally, which is slightly shifted by other contributions. There still could be a dominant fluorescent source present in these samples, but the corresponding peak would be outside our measured wavelength domain, see Theory 2.4.

### 5.2.3 Possibilities for application in mismatch studies

That AON1 to AON3 cannot be used is not problematic in the context of medical applications since there are many AON-designs possible, far more than treated in this thesis. When generalising the conclusion from this experiment, the location dependent mismatch sensitivity of 2/5 of AONs can be found using the AON-probe technique. This is still sufficient to provide insight into binding behaviour that can significantly contribute to AON development for exon skipping. Even so, optimising the procedure could allow us to apply the probe technique to the other AONs potentially as well.

## 5.3 Mismatch experiment

### 5.3.1 Normalisation and AON choice

The AON-testing results provide two suitable candidates for the mismatch experiment: AON4 and AON5. In a previous study, an AON with 23 bases was used [56]. AON4 has 20 bases without the probe, while AON5 also has 23. AON4 is shorter than what has been studied already, thus using it in our research provides a new insight into the binding behaviour of shorter AONs. Furthermore, in medical applications, shorter AONs are more commonly used, since long strands will bind frequently to a non-target area, which should be prevented. The first strategy is to normalise by intensity from the maximum efficiency, corresponding to the complementary strand. This way, we use the binding efficiency of a fully complementary AON to quantify an 'efficiency' of 1 and measure the binding efficiencies of the mismatched strands, which should be lower, relative to



Version of July 20, 2017– Created July 20, 2017 - 18:18

this number.

However, it was found that the intensity from mismatches at terminals was often higher and the intensity from the complementary was considered insufficiently precise to guarantee its accuracy, based on intermediate results. It was therefore not used to normalise the results.

Instead, the intensity from mismatch position 20 was chosen as normalisation for the final result and all sessions in which it was measured. This choice is based on the experience from previous studies that the binding efficiency related ratio at the ends is closest to 1. In these studies, the result is normalised by the complementary strand. When position 20 was not measured during a session, position 19 was used. This is considered to be the second best option for normalisation since it is also close to the 3′ end. Position 1 or 2 was not chosen for the reason that the result close to and at the 5′ end is less accurate because of the proximity of the mismatch to the fluorescent silver cluster.

### 5.3.2 The experimental result

The overall shape of the experimental result in figure 4.3 suggests that the AON containing a single mismatch binds most efficiently when this mismatch is located at either end of the strand. It decreases further towards the middle where it ”oscillates” about 0.6. This is plausible since a similar result has been found in a study on surface bound DNA [15]. Especially that the end positions correspond to a high efficiency is well in accordance with studies that explicitly aim at the location dependency [13, 15, 53, 56].

The result from positions 14 and 16 may pose problems since binding efficiency 1.0 lies within their error bars. Hence an unexpectedly high binding efficiency, thus a low mismatch sensitivity for binding, cannot be excluded and further experiments are be required to have an accurate measurement of these points. Furthermore, the result at location 19 is above 1.0. This point, however, can be considered to be particularly inaccurate as the corresponding error bar is far larger than that of the normalisation value at position 20, for example. If the results are improved with future experiments, this point is expected to be below 1.0. Additionally, the error bar at position 2 does not even include values below 1. This can be expected because of the proximity of the mismatch to the 19b-probe encapsulating

the silver cluster, which can lead to an increase in intensity that does not correspond to the binding efficiency. For this reason, the result at position 1, which would be even less accurate, is not processed for this result. Additionally, the model cannot calculate the ratio for this mismatch location either, as explained in Theory 2.3.3.

The errors in this result are significant and the dominant cause is assumed to be the instability of the measurement, which is the undesired time-dependency of the intensity. This means that the order of measuring the samples influences the obtained intensity. To prevent the deforming of the result, the order has been randomised from the third measurement session onwards as mentioned in the Methods 3.3. This does not reduce the spread in intensity results, however. It is expected that solving the instability problem will greatly improve the accuracy and precision of the method. The instability is considered to be the dominant factor because the human error is moderate, see section 5.5 and the device' measurement was precise. Leaving a sample in the fluorimeter and measuring again within a minute leads to an overlapping spectrum, which indicates its consistency. If the spread is still large in future results, it is recommended to obtain more measurements to limit the effect of outliers. This will improve the accuracy, which is especially required for mismatch positions 1, 2, 14, 16 and 19 and the complementary AON.

### 5.3.3 The model calculations

The model predicts the effect of a mismatch on the binding efficiency normalised by that of the complementary strand. The probe-AON method should relate this effect on binding efficiency directly to the relative fluorescence intensity from the samples, in order to compare the model with the experimental result meaningfully. As discussed in section 5.3.2, the measured results for certain points, in particular, are considered to be inaccurate, thus the validity of the model cannot be evaluated based on those points.

The model uses separate terms corresponding to the left and right side of the mismatch, thus the effect at end position 20 and 1 cannot be calculated, which explains the absence of this point.

In the middle range, especially from position 7 to 14, the model is physi-



Version of July 20, 2017– Created July 20, 2017 - 18:18

cally plausible since the points lie within the error bars of the experimental result. At some of these points, the difference is still significant: up to roughly 0.25. This difference seems to be proportional to the error leading to the author's hypothesis that the model is more accurate than the experiment in this range.

Close to the edges, we observe a notable difference between model-generated points and experimental ones. Due to the instability of the fluorescence, the author of this thesis does not exclude systematic errors causing this difference. Nevertheless, a previous study using the same method obtained similar results, where the experimental result is far below the model curve close to the edges, while these measurements were stable. Furthermore, nothing indicated that a structural error not related to instability has caused this difference and it is therefore considered to be a model inaccuracy.

An explanation of this deviation is the effect of dangling ends that is linear with the length of the dangling sections in this model. In reality, the effect increases rapidly with this length, thus the binding efficiency drops rapidly as well, with the mismatch 2 positions from the 3′ or 5′ end onwards, in both studies.

## 5.4 Stability of intensity measurement

### 5.4.1 Observations and hypothesis

From fig 4.4 in the Results one can conclude that the intensity is stable for the probe after approximately 20 minutes.Therefore, the mismatch measurements obtained later were expected to be stable. This is substantiated by the experience from the probe-AON strands in a previous study[56]. Based on all previous experience with fluorescent Ag-DNA, the intensity of all samples should level exponentially, just like the result from the probe. Because, in the result when the mismatch is located at position 14, the signal evolution initially has an increasing derivative, fitting the behaviour with a single exponential would be impossible. It should be possible to fit a single exponential to only the second half of this curve but this would lead to an arbitrary interpretation of the development which would be problematic as well. Applying the fit to the complementary AON strand does indicate the expected time dependency during stabilisation.

The assumption that the evolution of the signal over time follows a single exponential function is based on being able to approximate the formation of a fluorescent cluster as a first order chemical reaction. This should be the case when the only occurring process is the formation of fluorescent clusters from the available silver atoms and ions. This is only the case, however, if the other reactions, the reduction of the silver ions into atoms, and the binding of the silver ions to the DNA strands, have completed fully. The differences in the measured behaviour, especially in the early stages of the measurement, indicate that these reactions do not complete at the same rate for all samples, causing differences in their emission over time.

Both strands' intensities were not constant at the end of the stability measurements as deduced from the slope at the last point. This means that the intensity was not stable during the measurements of the mismatched samples that started just after the stability measurement. The time development was not monitored during these experiments, thus it is not known in that period. Although the intensity could have become constant for several minutes at some point, the instability at the start still has a significant impact on the result. It is also plausible that, before the last sample was put in the fluorimeter, the intensity had already started to decrease, leading to an even greater difference in intensity between samples.

Since the method relies on time independence of the emission throughout the measurement this strongly suggests that the instability is a significant factor influencing the mismatch experiment result.

### 5.4.2 Implications

The stability data was processed after the sessions, so this provides only a hindsight check of stability, thus no corrections could be made to the experiments. This problem could not be solved in the limited time available for the BSc research and instead, the author chose to continue experiments with the longer stabilising time of an hour to obtain an insightful result nevertheless.

During the mismatch measurements, which were conducted right after the stability measurement, the time development of each sample's inten-

sity is uncertain. From previous experience with fluorescent Ag-DNA, it is known that the intensity decreases after several hours or earlier [56]. Therefore it is assumed for that from all samples in this research the emitted intensity eventually falls as well.

From the results, the conclusion is drawn that the time between adding the reductor and measuring the samples should be prolonged in future experiments. Furthermore, the stability should be ensured before measuring any other samples. Additionally, the time-dependent behaviour should be studied over longer periods to evaluate its consistency. The purpose of this is to properly schedule the measurement period in the interval that the intensity is stable or to be able to correct for the time development of the signal. If it is not consistent other measures should be taken to improve the accuracy of the results, such as a simultaneous measurement of a control-sample to be able to correct for instabilities or simply obtaining far more measurement points.

## 5.5 Error involved in the preparation of the samples

Compared to the total error in the results, the error involved in the sample preparation is relatively small: about 0.13. If this error is considered as an independent addition, the total error is $\sqrt{\sigma^2 + 0.13^2}$ in which $\sigma$ is the standard deviation due to other influences. Since the total error is above 0.20 for most points, the contribution of the preparation error is small. Let us, for example, take $\sigma$ as 0.3, then the total error will be$\sqrt{0.3^2 + 0.13^2} \approx 0.33$, while it would be just 0.3 if the preparation error was absent. Therefore the total error increases by only 0.03, which indicates an insignificant impact. One can deduce or determine by filing in higher values for $\sigma$ that the contribution decreases for increasing $\sigma$ (or total error). That said, the measurement puts a minimum on the error in the used Ag-DNA probe method of 0.13 if no other effects were present. This number is comparable to the error observed in previous measurements on a different AON [56].

Chapter 6

# Conclusion

The developments to treat Duchenne muscular dystrophy are advancing. A cure, however, is not expected to be found in the near future. An approach called ”exon skipping” uses molecules that bind to specific sites in the genetic material to slow the progression of this impactful disease. This is not safe when the drug components alter the expression of ”healthy” parts of the genetic code as well. An optimal exclusivity of the binding is required to prevent this. This can only be achieved with a substantially improved understanding of the binding behaviour of nucleotide strands, i.e. organic RNA and DNA in the context of this thesis.

An important factor is the presence of mismatches, i.e. defects that inhibit the formation of base pairs. In particular, the location of such a mismatch in a base sequence has a significant effect on binding efficiency, thus a novel model has been developed for this purpose.

To evaluate this model, an experimental result is required under physiological conditions in which the location dependency is isolated from other factors. This BSc research provides this experiment and compares the model to the obtained result. It uses a silver nanocluster encapsulated in a DNA probe to relate the emission from the silver to the binding efficiency.

Both the experimental result and the model calculation are deemed physically plausible and in accordance with each other for most of the mismatch positions. The similarity to results from other studies is clear, which contributes to this conclusion. However, the methodological differences between those studies and this BSc research are likely to cause the result to

be different as well.

Although the found relation is considered to be generally valid, we observe some significant inaccuracies, in particular when the mismatch is close to the ends of the strands. Depending on future results, the model may need to include additional parameters to find better agreement with the experimental data in these positions.

Additionally, the error in the measurement was substantial. This is probably caused by the instability of the fluorescence of the probe. Furthermore, the sensitivity of the technique is dependent on the mismatch position relative to the probe. Therefore, in the future, further care must be taken in ensuring the stability of the fluorescence during measurement, and in positioning the probe on either end of the AON strand.

Despite this, the probe technique itself provides a promising approach for measuring the binding efficiency of nucleotide strands due to its simplicity, low cost, short duration of the experiment and high resource availability.

Chapter 7

# Acknowledgements

First and foremost I would like to thank Donny de Bruin Bsc for the great effort, time and academic guidance with which he supervised me during this project. His help has been more than I allow myself to expect. I thank him and Nelli Bossert MSc for offering this interesting project and for providing the information in the introductory period after which I chose this project. I thank Nelli also for her willingness to review a concept version of my thesis and her interest in the progress of my research. I thank Prof.dr. Dirk Bouwmeester for his offer to review my thesis and his help to decide which project to choose. I thank Dr. Chris Smiet and Dr. Vasco Tenner for their occasional help and interesting discussions. I thank the LION [1] Quantum Optics group for their hospitality, advice, help and discussions. From those, I would like to thank Dr. Wolfgang Löffler for his profound concern with, and interest in all group members, students and staff alike. I am also grateful for his help when I encountered problems during the editing of my thesis. I thank Ing. Marcel Winter for his advice on laboratory work. I am grateful for my parents who made it possible for me to study successfully, advised me, reviewed my work and helped me manage this project. Furthermore, I have been inspired by Prof.dr.ir. Tjerk Oosterkamp's lecture on managing one's responsibility as a scientist and coping with disappointments in research. I thank the internet for its dank memes, top kek and YTPs when I needed some distraction. Lastly, I thank the scientific community for these interesting reads [77–81].

[1] Leids Instituut voor Onderzoek in Natuurkunde

Chapter 8

# Appendix

## 8.1 Acronyms and definitions

**19b-probe** Ag-DNA probe with a clear fluorescence spectrum which is added to other strands to link emitted intensity to binding efficiency.

**A** Adenine, see ”Nucleotide”

**AON** ”Antisense oligonucleotide. Small pieces of DNA or RNA that can bind to specific molecules of RNA. This blocks the ability of the RNA to make a protein or work in other ways. Antisense oligonucleotides may be used to block the production of proteins needed for cell growth. They are being studied in the treatment of several types of cancer. Also called antisense agent.” [82]

**Base** See ”Nucleotide”.

**BMD** Becker Muscular Dystrophy. A muscular dystrophy similar to DMD, however less severe. exon skipping treatments can achieve the DMD to be effectively BMD in patients.

**C** Cytosine, see ”Nucleotide”.

**Complementary** Able to form a perfectly matching double-stranded structure with the opposite nucleotide strand.

**DMD** ”Duchenne muscular dystrophy: The best-known form of muscular dystrophy, due to mutation in a gene on the X chromosome that prevents the production of dystrophin, a normal protein in muscle.”[83]

**DNA** Abbreviation of deoxyribonucleic acid. A double-stranded nucleic acid that contains the genetic information for cell growth, division, and function. [84]

**DNA/DNA** See "X/Y"

**Duplex** "Duplex DNA is another name for double-stranded DNA. This means that the nucleotides of two DNA sequences have bonded together and then coiled to form a double helix. This double-stranded structure facilitates the stable duplication of genetic material, a requirement for cell division."[85]

**Dystrophin** "A protein located mainly in the heart and skeletal muscles involved in strengthening of muscle fibers preventing from injury as muscle contract and relax. Dystrophin gene is mapped on chromosome Xp21.2 containing 86 exons that span at about 2,300kb with 3026-3345 amino acids having ZZ domain which is responsible for Duchenne and Becker muscular dystrophies. Dystrophin serves as an anchor that connects each muscle cells framework which plays an important role in signaling through interacting with proteins that send and receive chemical signals. It is expressed at higher level in neuronal cells than astroglial cells in adult brain in which overexpression of this protein prevents the growth of abnormal mechanical properties linked with dystrophic muscle." [84]

**Exon** "Genes contain exons which are regions coding for proteins and which are interrupted by the unused sequences called introns. Exons have been found to include both sequences coding for amino acids and untranslated sequences. The introns are removed and the exons are joined together to form the final functional mRNA."[84]

**Exon skipping** The a treatment currently in development for Duchenne muscular dystrophy (DMD), where AONs are used to allow the synthesis of partly functional dystrophin proteins instead of non-functional ones. [86]

**G** Guanine, see "Nucleotide"

**Hybridisation** The binding of single strands to opposite strands to form a double stranded complex (a duplex). This occurs at a sufficiently low temperature.

**Hydrogen bond** The hydrogen bond is really a special case of dipole forces. A hydrogen bond is the attractive force between the hydrogen attached to an electronegative atom of one molecule and an electronegative atom of a different molecule. Usually the electronegative atom is oxygen, nitrogen, or fluorine, which has a partial negative charge. The hydrogen then has the partial positive charge.[29]

**Intron** See "Exon".

**Melting temperature ($T_M$)** "That temperature at which, under a given set of conditions, double-stranded dNA is changed (50%) to single-stranded DNA; under standard conditions, the base composition of the DNA can be estimated from the denaturation temperature, since the greater the denaturation temperature, the greater the guanine-plus-cytosine content (i.e., GC content) of the DNA. "[84]

**Mismatch** A defect in a nucleotide strand with an obnormal hydrogen bond pattern inhibiting the formation of a base pair, e.g. a guanine opposite of a thymine base.

**mRNA** "Messenger RNA (mRNA) is a subtype of RNA. An mRNA molecule carries a portion of the DNA code to other parts of the cell for processing. mRNA is created during transcription. During the transcription process, a single strand of DNA is decoded by RNA polymerase, and mRNA is synthesized. Physically, mRNA is a strand of nucleotides known as ribonucleic acid, and is single-stranded. "[87]

**Nearest-Neighbour (NN) model** A model to calculate the binding energy of a strand by using sequence dependent parameters, one for each nucleobase and its adjacent base. The traditional version of this model can only be applied to perfectly matching strands and that is the only well established one.

**NN** See "Nearest-Neighbour model"

**(Nucleo)base** Coding component of nucleotide strands that form hydrogen bonds with opposite bases, forming a base pair. The common types are A, G, C and T (or U), mentioned in the definition of "Nucleotide".

**Nucleotide** "The basic building block of nucleic acids, such as DNA and RNA. It is an organic compound made up of nitrogenous base, a sugar, and a phosphate group. DNA molecule consists of nucleotides

in which the sugar component is deoxyribose whereas the RNA molecule has nucleotides in which the sugar is a ribose. The most common nucleotides are divided into purines and pyrimidines based on the structure of the nitrogenous base. In DNA, the purine bases include adenine and guanine while the pyrimidine bases are thymine and cytosine. RNA includes adenine, guanine, cytosine, and uracil in stead of thymine (thymine is produced by adding a methyl to uracil).”[84]

**Pre-mRNA** ”An immature or incompletely processed mRNA molecule in eukaryotes originated from primary transcript (hnRNA) in the nucleus and need to be processed before it becomes a fully functional mature mRNA for transport into the cytoplasm.” [84]

**Purines** See ”Nucleotide”

**Pyrimidines** See ”Nucleotide”

**RNA** ”Ribonucleic acid (RNA) is a linear molecule composed of four types of smaller molecules called ribonucleotide bases: adenine (A), cytosine (C), guanine (G), and uracil (U). RNA is often compared to a copy from a reference book, or a template, because it carries the same information as its DNA template but is not used for long-term storage.”[87]

**Splicing** ”Removal of introns and connecting of exons in eukaryotic pre-mRNAs to form the mRNA.”[87]

**T** Thymine, see ”Nucleotide”

**X/Y** The slash means in biological context ”Nucleotide strand X bound to the opposite strand Y”, for example DNA/RNA, or AON/Exon

## 8.2 Used DNA sequences

### 8.2.1 Variable sequences

| Name | Sequence |
|---|---|
| TestAON1 | TGCCTTTTGGGGACGGATACCTCTGTGATTTTATAACTT |
| TestAON2 | TGCCTTTTGGGGACGGATACCCACCATCACCCTCTGTGA |
| TestAON3 | TGCCTTTTGGGGACGGATACCCACCATCACCCTCTGTGATTT |
| TestAON4 | TGCCTTTTGGGGACGGATAGGTCACCCACCATCACCCTC |
| TestAON5 | TGCCTTTTGGGGACGGATACCTCAAGGTCACCCACCATCACC |
| | |
| TestAON4-MM1 | TGCCTTTTGGGGACGGATATGTCACCCACCATCACCCTC |
| TestAON4-MM2 | TGCCTTTTGGGGACGGATAGTTCACCCACCATCACCCTC |
| TestAON4-MM3 | TGCCTTTTGGGGACGGATAGGACACCCACCATCACCCTC |
| TestAON4-MM4 | TGCCTTTTGGGGACGGATAGGTTACCCACCATCACCCTC |
| TestAON4-MM5 | TGCCTTTTGGGGACGGATAGGTCTCCCACCATCACCCTC |
| TestAON4-MM6 | TGCCTTTTGGGGACGGATAGGTCATCCACCATCACCCTC |
| TestAON4-MM7 | TGCCTTTTGGGGACGGATAGGTCACTCACCATCACCCTC |
| TestAON4-MM8 | TGCCTTTTGGGGACGGATAGGTCACCTACCATCACCCTC |
| TestAON4-MM9 | TGCCTTTTGGGGACGGATAGGTCACCCTCCATCACCCTC |
| TestAON4-MM10 | TGCCTTTTGGGGACGGATAGGTCACCCATCATCACCCTC |
| TestAON4-MM11 | TGCCTTTTGGGGACGGATAGGTCACCCACTATCACCCTC |
| TestAON4-MM12 | TGCCTTTTGGGGACGGATAGGTCACCCACCTTCACCCTC |
| TestAON4-MM13 | TGCCTTTTGGGGACGGATAGGTCACCCACCAACACCCTC |
| TestAON4-MM14 | TGCCTTTTGGGGACGGATAGGTCACCCACCATTACCCTC |
| TestAON4-MM15 | TGCCTTTTGGGGACGGATAGGTCACCCACCATCTCCCTC |
| TestAON4-MM16 | TGCCTTTTGGGGACGGATAGGTCACCCACCATCATCCTC |
| TestAON4-MM17 | TGCCTTTTGGGGACGGATAGGTCACCCACCATCACTCTC |
| TestAON4-MM18 | TGCCTTTTGGGGACGGATAGGTCACCCACCATCACCTTC |
| TestAON4-MM19 | TGCCTTTTGGGGACGGATAGGTCACCCACCATCACCCAC |
| TestAON4-MM20 | TGCCTTTTGGGGACGGATAGGTCACCCACCATCACCCTT |

### 8.2.2 Target sequence

Exon51-loc28-29: TGGCTTTCTCTGCTTGATCAAGTTATAAAATCACA-
GAGGGTGATGGTGGGTGACCTTGAGGATATCAACGAGATGATCATCAAGCAGAAG

## 8.3 Recipe for AON-testing experiment

1. Insert 60 $\mu l$ of the 25 mM HEPES-NaOH pH 7.4 buffer into the 1.5 ml tube.

2. Add 20 $\mu$l of the 100 $\mu$M Exon DNA solution for the bound configuration or 20 $\mu$l nuclease-free water for the unbound configuration.

3. Add 20 $\mu$l of the 100 $\mu$M AON DNA solution. After this step the tubes are centrifuged for 10 seconds at 10,000 rpm to mix their contents properly and to avoid droplets on the walls[1].

4. Keep all tubes at 37 °C for two hours.

5. Add 10 $\mu$l of the 1.92 mM $AgNO_3$ solution.

6. Keep all tubes at 37 °C in the dark for 30 minutes and prepare a $NaBH_4$ solution of 0.96 mM.

7. Add 10 $\mu$l 0.96 mM $NaBH_4$ solution.

[1] This measure was only taken during some of the mismatch experiments described in section Mismatch experiment.

# Bibliography


[1] E. P. Hoffman, A. Bronson, A. A. Levin, S. Takeda, T. Yokota, A. R. Baudy, and E. M. Connor, *Restoring Dystrophin Expression in Duchenne Muscular Dystrophy Muscle*, The American Journal of Pathology **179**, 12 (2011).

[2] J. SantaLucia, *A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics*, Proceedings of the National Academy of Sciences **95**, 1460 (1998).

[3] E. G. Gwinn and D. Bouwmeester, *DNA nanophotonics using artificial neural networks to analyze fluorescent Ag:DNA.*, Unpublished .

[4] P. Melacini, A. Vianello, C. Villanova, M. Fanin, M. Miorin, C. Angelini, and S. Dalla Volta, *Cardiac and respiratory involvement in advanced stage Duchenne muscular dystrophy*, Neuromuscular Disorders **6**, 367 (1996).

[5] M. Yoshida and E. Ozawa, *Glycoprotein Complex Anchoring Dystrophin to Sarcolemma*, The Journal of Biochemistry **108**, 748 (1990).

[6] A. Nakamura, *Moving towards successful exon-skipping therapy for Duchenne muscular dystrophy*, Journal of Human Genetics (2017).

[7] A. Aartsma-Rus and A. M. Krieg, *FDA Approves Eteplirsen for Duchenne Muscular Dystrophy: The Next Chapter in the Eteplirsen Saga*, Nucleic Acid Therapeutics **27**, 1 (2016).

[8] J. Hooyberghs, P. Van Hummelen, and E. Carlon, *The effects of mismatches on hybridization in DNA microarrays: determination of nearest neighbor parameters*, Nucleic Acids Research **37**, e53 (2009).

[9] J. SantaLucia, H. T. Allawi, and P. A. Seneviratne, *Improved nearest-neighbor parameters for predicting DNA duplex stability*, Biochemistry **35**, 3555 (1996).

[10] N. Sugimoto, S. Nakano, M. Yoneyama, and K. Honda, *Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes.*, Nucleic Acids Research **24**, 4501 (1996).

[11] J. W. Nelson, F. H. Martin, and I. Tinoco, *DNA and RNA oligomer thermodynamics: The effect of mismatched bases on double-helix stability*, Biopolymers **20**, 2509 (1981).

[12] E. H. Niks and A. Aartsma-Rus, *Exon skipping: a first in class strategy for Duchenne muscular dystrophy*, Expert Opinion on Biological Therapy **17**, 225 (2017).

[13] T. Naiser, O. Ehler, J. Kayser, T. Mai, W. Michel, and A. Ott, *Impact of point-mutations on the hybridization affinity of surface-bound DNA/DNA and RNA/DNA oligonucleotide-duplexes: Comparison of single base mismatches and base bulges*, BMC Biotechnology **8**, 48 (2008).

[14] A. Pozhitkov, P. A. Noble, T. Domazet-Lošo, A. W. Nolte, R. Sonnenberg, P. Staehler, M. Beier, and D. Tautz, *Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted*, Nucleic Acids Research **34**, e66 (2006).

[15] T. Naiser, J. Kayser, T. Mai, W. Michel, and A. Ott, *Position dependent mismatch discrimination on DNA microarrays - experiments and model*, BMC Bioinformatics **9**, 509 (2008).

[16] C. Rennie, H. A. Noyes, S. J. Kemp, H. Hulme, A. Brass, and D. C. Hoyle, *Strong position-dependent effects of sequence mismatches on signal ratios measured using long oligonucleotide microarrays*, BMC Genomics **9**, 317 (2008).

[17] A. Aartsma-Rus, L. van Vliet, M. Hirschi, A. A. M. Janson, H. Heemskerk, C. L. de Winter, S. de Kimpe, J. C. T. van Deutekom, P. A. C. 't Hoen, and G.-J. B. van Ommen, *Guidelines for antisense oligonucleotide design and insight into splice-modulating mechanisms.*, Molecular therapy : the journal of the American Society of Gene Therapy **17**, 548 (2009).

[18] L. PASSAMANO, A. TAGLIA, A. PALLADINO, E. VIGGIANO, P. D'AMBROSIO, M. SCUTIFERO, M. ROSARIA CECIO, V. TORRE, F. DE LUCA, E. PICILLO, O. PACIELLO, G. PILUSO, G. NIGRO, and L. POLITANO, *Improvement of survival in Duchenne Muscular Dystrophy: retrospective analysis of 835 patients*, Acta Myologica **31**, 121 (2012).

[19] E. P. Hoffman, R. H. Brown, and L. M. Kunkel, *Dystrophin: The protein product of the duchenne muscular dystrophy locus*, Cell **51**, 919 (1987).

[20] J. R. Mendell, C. Shilling, N. D. Leslie, K. M. Flanigan, R. al Dahhak, J. Gastier-Foster, K. Kneile, D. M. Dunn, B. Duval, A. Aoyagi, C. Hamil, M. Mahmoud, K. Roush, L. Bird, C. Rankin, H. Lilly, N. Street, R. Chandrasekar, and R. B. Weiss, *Evidence-based path to newborn screening for duchenne muscular dystrophy*, Annals of Neurology **71**, 304 (2012).

[21] R. R. Kaprielian, S. Stevenson, S. M. Rothery, M. J. Cullen, and N. J. Severs, *Distinct Patterns of Dystrophin Organization in Myocyte Sarcolemma and Transverse Tubules of Normal and Diseased Human Myocardium*, Circulation **101**, 2586 (2000).

[22] A. P. Monaco, C. J. Bertelson, S. Liechti-Gallati, H. Moser, and L. M. Kunkel, *An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus*, Genomics **2**, 90 (1988).

[23] W. Gilbert, *Genes-in-pieces revisited*, Science **228**, 823 (1985).

[24] G. Streisinger, Y. Okada, J. Emrich, J. Newton, A. Tsugita, E. Terzaghi, and M. Inouye, *Frameshift Mutations and the Genetic Code*, Cold Spring Harbor Symposia on Quantitative Biology **31**, 77 (1966).

[25] V. Arechavala-Gomeza, I. Graham, L. Popplewell, A. Adams, A. Aartsma-Rus, M. Kinali, J. Morgan, J. van Deutekom, S. Wilton, G. Dickson, and F. Muntoni, *Comparative Analysis of Antisense Oligonucleotide Sequences for Targeted Skipping of Exon 51 During Dystrophin Pre-mRNA Splicing in Human Muscle*, Human Gene Therapy **18**, 798 (2007).

[26] A. Aartsma-Rus, M. Bremmer-Bout, A. A. M. Janson, J. T. Den Dunnen, G. J. B. Van Ommen, and J. C. T. Van Deutekom, *Targeted exon skipping as a potential gene correction therapy for Duchenne muscular dystrophy*, Neuromuscular Disorders **12** (2002).

[27] S. Jirka and A. Aartsma-Rus, *An update on RNA-targeting therapies for neuromuscular disorders:*, Current Opinion in Neurology **28**, 515 (2015).

[28] T.-b. Zhang, C.-l. Zhang, Z.-l. Dong, and Y.-f. Guan, *Determination of Base Binding Strength and Base Stacking Interaction of DNA Duplex Using Atomic Force Microscope*, Scientific Reports **5** (2015).

[29] E. Arunan, G. R. Desiraju, R. A. Klein, J. Sadlej, S. Scheiner, I. Alkorta, D. C. Clary, R. H. Crabtree, J. J. Dannenberg, P. Hobza, H. G. Kjaergaard, A. C. Legon, B. Mennucci, and D. J. Nesbitt, *Definition of the hydrogen bond (IUPAC Recommendations 2011)*, Pure and Applied Chemistry **83**, 1637 (2011).

[30] M. Mandel and J. Marmur, *Use of ultraviolet absorbance-temperature profile for determining the guanine plus cytosine content of DNA*, Methods in Enzymology **12**, 195 (1968).

[31] P. Doty, H. Boedtker, J. R. Fresco, B. D. Hall, and R. Haselkorn, *Configurational Studies of Polynucleotides and Ribonucleic Acid**, Annals of the New York Academy of Sciences **81**, 693 (1959).

[32] F.-C. Chou, W. Kladwang, K. Kappel, and R. Das, *Blind tests of RNA nearest-neighbor energy prediction*, Proceedings of the National Academy of Sciences **113**, 8430 (2016).

[33] J. Zhu and R. M. Wartell, *The Relative Stabilities of Base Pair Stacking Interactions and Single Mismatches in Long RNA Measured by Temperature Gradient Gel Electrophoresis*, Biochemistry **36**, 15326 (1997).

[34] N. E. Watkins, W. J. Kennelly, M. J. Tsay, A. Tuin, L. Swenson, H.-R. Lee, S. Morosyuk, D. A. Hicks, and J. SantaLucia, *Thermodynamic contributions of single internal rAÂ·dA, rCÂ·dC, rGÂ·dG and rUÂ·dT mismatches in RNA/DNA duplexes*, Nucleic Acids Research **39**, 1894 (2011).

[35] H. T. Allawi and J. SantaLucia, *Nearest Neighbor Thermodynamic Parameters for Internal G-A Mismatches in DNA*, Biochemistry **37**, 2170 (1998).

[36] H. T. Allawi and J. SantaLucia, *Nearest-Neighbor Thermodynamics of Internal A-C Mismatches in DNA: Sequence Dependence and pH Effects*, Biochemistry **37**, 9435 (1998).

[37] C. Trapp, M. Schenkelberger, and A. Ott, *Stability of double-stranded oligonucleotide DNA with a bulged loop: a microarray study*, BMC Biophysics **4**, 20 (2011).

[38] J. John SantaLucia and D. Hicks, *The Thermodynamics of DNA Structural Motifs*, Annual Review of Biophysics and Biomolecular Structure **33**, 415 (2004).

[39] M. J. McCauley, C. J. Cain, L. Furman, C. A. Dietrich, S. Ruderman, D. Seminario-McCormick, G. Ferris, M. E. Nunez, and M. C. Williams, *Kinetics and Thermodynamics of Non-Canonical DNA*, Biophysical Journal **110**, 407a (2016).

[40] K. J. Luebke, R. P. Balog, and H. R. Garner, *Prioritized selection of oligodeoxyribonucleotide probes for efficient hybridization to RNA transcripts*, Nucleic Acids Research **31**, 750 (2003).

[41] L. Zhang, M. F. Miles, and K. D. Aldape, *A model of molecular interactions on short oligonucleotide microarrays*, Nature Biotechnology **21**, 818 (2003).

[42] E. Carlon and T. Heim, *Thermodynamics of RNA/DNA hybridization in high-density oligonucleotide microarrays*, Physica A: Statistical Mechanics and its Applications **362**, 433 (2006).

[43] H. Urakawa, S. E. Fantroussi, H. Smidt, J. C. Smoot, E. H. Tribou, J. J. Kelly, P. A. Noble, and D. A. Stahl, *Optimization of Single-Base-Pair Mismatch Discrimination in Oligonucleotide Microarrays*, Applied and Environmental Microbiology **69**, 2848 (2003).

[44] P. H. Hagedorn, B. R. Hansen, T. Koch, and M. Lindow, *Managing the sequence-specificity of antisense oligonucleotides in drug discovery*, Nucleic Acids Research **45**, 2262 (2017).

[45] J.-L. Ma, B.-C. Yin, H.-N. Le, and B.-C. Ye, *Label-Free Detection of Sequence-Specific DNA Based on Fluorescent Silver Nanoclusters-Assisted Surface Plasmon-Enhanced Energy Transfer*, ACS Applied Materials & Interfaces **7**, 12856 (2015).

[46] A. Granzhan, N. Kotera, and M.-P. Teulade-Fichou, *Finding needles in a basestack: recognition of mismatched base pairs in DNA by small molecules*, Chemical Society Reviews **43**, 3630 (2014).

[47] M. Schenkelberger, C. Trapp, T. Mai, and A. Ott, *Ultrahigh molecular recognition specificity of competing DNA oligonucleotide strands in thermal equilibrium*, (2016).

[48] S. Torgasin and K.-H. Zimmermann, *Algorithm for thermodynamically based prediction of DNA/DNA cross-hybridisation*, International Journal of Bioinformatics Research and Applications **6**, 82 (2010).

[49] R. Kierzek, M. E. Burkard, and D. H. Turner, *Thermodynamics of Single Mismatches in RNA Duplexes*, Biochemistry **38**, 14214 (1999).

[50] J. Wang, *Electrochemical nucleic acid biosensors*, Analytica Chimica Acta **469**, 63 (2002).

[51] A. R. Davis and B. M. Znosko, *Positional and Neighboring Base Pair Effects on the Thermodynamic Stability of RNA Single Mismatches*, Biochemistry **49**, 8669 (2010).

[52] F. Duan, M. A. Pauley, E. R. Spindel, L. Zhang, and R. B. Norgren, *Large scale analysis of positional effects of single-base mismatches on microarray gene expression data*, BioData Mining **3**, 2 (2010).

[53] T. Naiser, J. Kayser, T. Mai, W. Michel, and A. Ott, *Stability of a Surface-Bound Oligonucleotide Duplex Inferred from Molecular Dynamics: A Study of Single Nucleotide Defects Using DNA Microarrays*, Physical Review Letters **102**, 218301 (2009).

[54] S. Suzuki, N. Ono, C. Furusawa, A. Kashiwagi, and T. Yomo, *Experimental optimization of probe length to increase the sequence specificity of high-density oligonucleotide microarrays*, BMC Genomics **8**, 373 (2007).

[55] M. Nazmul Alam, M. Hassan Shamsi, and H.-B. Kraatz, *Scanning positional variations in single- nucleotide polymorphism of DNA: an electrochemical study*, Analyst **137**, 4220 (2012).

[56] D. de Bruin and D. Bouwmeester, *Studying DNA hybridization using fluorescent DNA-stabilized silver clusters to investigate nucleotide-mismatch dependencies*, Unpublished (2017).

[57] T. Vosch, Y. Antoku, J.-C. Hsiang, C. I. Richards, J. I. Gonzalez, and R. M. Dickson, *Strongly emissive individual DNA-encapsulated Ag nanoclusters as single-molecule fluorophores*, Proceedings of the National Academy of Sciences **104**, 12616 (2007).

[58] A. W. Peterson, L. K. Wolf, and R. M. Georgiadis, *Hybridization of Mismatched or Partially Matched DNA at Surfaces*, Journal of the American Chemical Society **124**, 14601 (2002).

[59] F. F. Millenaar, J. Okyere, S. T. May, M. van Zanten, L. A. Voesenek, and A. J. Peeters, *How to decide? Different methods of calculating gene expression from short oligonucleotide array data will give different results*, BMC Bioinformatics **7**, 137 (2006).

[60] L. Zhang, C. Wu, R. Carta, and H. Zhao, *Free energy of DNA duplex formation on short oligonucleotide microarrays*, Nucleic Acids Research **35**, e18 (2007).

[61] H. Binder, *Thermodynamics of competitive surface adsorption on DNA microarrays*, Journal of Physics: Condensed Matter **18**, S491 (2006).

[62] S. M. Copp, D. Schultz, S. M. Swasey, A. Faris, and E. G. Gwinn, *Cluster Plasmonics: Dielectric and Shape Effects on DNA-Stabilized Silver Clusters*, Nano Letters **16**, 3594 (2016).

[63] N. Bossert et al., *Fluorescence-tunable Ag-DNA biosensor with tailored cytotoxicity for live-cell applications*, Scientific Reports **6**, 37897 (2016).

[64] P. R. O'Neill, L. R. Velazquez, D. G. Dunn, E. G. Gwinn, and D. K. Fygenson, *Hairpins with Poly-C Loops Stabilize Four Types of Fluorescent Agn:DNA*, The Journal of Physical Chemistry C **113**, 4229 (2009).

[65] J. Sharma, H.-C. Yeh, H. Yoo, J. H. Werner, and J. S. Martinez, *A complementary palette of fluorescent silver nanoclusters*, Chemical Communications **46**, 3280 (2010).

[66] E. G. Gwinn, P. O'Neill, A. J. Guerrero, D. Bouwmeester, and D. K. Fygenson, *Sequence-dependent fluorescence of DNA-hosted silver nanoclusters*, Advanced Materials **20**, 279 (2008).

[67] J. T. Petty, J. Zheng, N. V. Hud, and R. M. Dickson, *DNA-Templated Ag Nanocluster Formation*, Journal of the American Chemical Society **126**, 5207 (2004).

[68] C. I. Richards, S. Choi, J.-C. Hsiang, Y. Antoku, T. Vosch, A. Bongiorno, Y.-L. Tzeng, and R. M. Dickson, *Oligonucleotide-Stabilized Ag Nanocluster Fluorophores*, Journal of the American Chemical Society **130**, 5038 (2008).

[69] D. Schultz and E. G. Gwinn, *Silver atom and strand numbers in fluorescent and dark Ag:DNAs.*, Chemical communications (Cambridge, England) **48**, 5748 (2012).

[70] W. Guo, J. Yuan, Q. Dong, and E. Wang, *Highly Sequence-Dependent Formation of Fluorescent Silver Nanoclusters in Hybridized DNA Duplexes for Single Nucleotide Mutation Identification*, Journal of the American Chemical Society **132**, 932 (2010).

[71] Z. Huang, F. Pu, D. Hu, C. Wang, J. Ren, and X. Qu, *Site-Specific DNA-Programmed Growth of Fluorescent and Functional Silver Nanoclusters*, Chemistry - A European Journal **17**, 3774 (2011).

[72] K. Ma, Q. Cui, G. Liu, F. Wu, S. Xu, and Y. Shao, *DNA abasic site-directed formation of fluorescent silver nanoclusters for selective nucleobase recognition*, Nanotechnology **22**, 305502 (2011).

[73] A. Latorre and A. Somoza, *DNA-Mediated Silver Nanoclusters: Synthesis, Properties and Applications*, ChemBioChem **13**, 951 (2012).

[74] S. M. Copp, D. Bouwmeester, and E. Gwinn, *Optical Materials with a Genome: Nanophotonics with DNA-Stabilized Silver Clusters*, PhD thesis, University of California Santa Barbara, Santa Barbara, US, 2016.

[75] M. Berdakin, M. Taccone, K. J. Julian, G. Pino, and C. G. Sánchez, *Disentangling the Photophysics of DNA-Stabilized Silver Nanocluster Emitters*, The Journal of Physical Chemistry C **120**, 24409 (2016).

[76] D. Schultz and E. Gwinn, *Stabilization of fluorescent silver clusters by RNA homopolymers and their DNA analogs: C,G versus A,T(U) dichotomy.*, Chemical communications **47**, 4715 (2011).

[77] R. N. Salaman and W. G. Burton, *The History and Social Influence of the Potato*, Cambridge University Press, 1985, Google-Books-ID: EV4YE_0RsywC.

[78] G. Hine, J. Onaolapo, E. De Cristofaro, N. Kourtellis, I. Leontiadis, R. Samaras, G. Stringhini, and J. Blackburn, *Kek, Cucks, and God Emperor Trump: A Measurement Study of 4chan's Politically Incorrect Forum and its Effects on the Web*, 2017.

[79] L. J. Powell and E. S. D. Engelhardt, *The Perilous Whiteness of Pumpkins*, GeoHumanities **1**, 414 (2015).

[80] J. Bowling and B. Martin, *Science: a masculine disorder?*, Science and Public Policy **12**, 308 (1985).

[81] F. F. Morin, *Ego Hippo*, Angelaki **22**, 87 (2017).

[82] *NCI Dictionary of Cancer Terms*.

[83] *Medical Definition of Duchenne muscular dystrophy*.

[84] *Biology-Online Dictionary*.

[85] *What Is Duplex DNA?*

[86] A. Aartsma-Rus, *Antisense-mediated modulation of splicing: therapeutic implications for Duchenne muscular dystrophy*, RNA Biol **7**, 453 (2010).

[87] *Scitable | Learn Science at Nature*.